UNIVERSITY OF COPENHAGEN

Scholarly communication changing

The Implications of open access

Faber Frandsen, Tove

Publication date:
2009

Document version
Early version, also known as pre-print

Citation for published version (APA):
Faber Frandsen, T. (2009). *Scholarly communication changing: The Implications of open access*. Danmarks BIblioteksskole.

Download date: 08. apr.. 2020

Scholarly communication changing:
The implications of open access

Tove Faber Frandsen

PhD thesis from the Research Programme of Knowledge Culture and Knowledge Media,
Royal School of Library and Information Science, Denmark

Scholarly communication changing:
The implications of open access

Tove Faber Frandsen

PhD thesis from the Research Programme of Knowledge Culture and Knowledge Media,
Royal School of Library and Information Science, Denmark

CIP – Cataloguing in Publication

Frandsen, Tove Faber

Scholarly communication changing: The implications of open access / Tove Faber Frandsen. – Copenhagen, Denmark: Royal School of Library and Information Science, 2009. – VIII, 272.
ISBN 9788774153122

ISBN 9788774153122

© Copyright Tove Faber Frandsen 2009

All rights reserved

Videnskabelig kommunikation i forandring: Implikationerne af open access

Tove Faber Frandsen

Ph.d.-afhandling fra Forskningsprogram for Videnskultur og vidensmedier,
Danmarks Biblioteksskole

Acknowledgements

It is a pleasure to thank all the people who have helped and inspired me during my doctoral study.

I would like to state my gratitude to my PhD supervisors, professor Birger Hjørland and associate professor Jeppe Nicolaisen. I gratefully acknowledge the support of both of these great scholars. You have provided the knowledgeable and stimulating discussions needed to challenge my work. I also thank you for being there when good friends were needed.

I would also like to thank Denmark's Electronic Research Library, particularly Hanne Marie Kværndrup and Lise Mikkelsen, for opening doors for me to the Danish open access community and for your enthusiastic involvement in this project.

This research also benefited tremendously from many colleagues at the Royal School of Library and Information Science. I would like to express my gratitude to all the members of the research programme, Knowledge Culture and Knowledge Media, for your warm welcome. I value your insightful comments and the friendly atmosphere in which we can discuss research. I am especially grateful to Gitte, Jack, Laura, Lise, Mikkel and Nanna. I would also like to thank the library and the research administration for assisting me in many different ways. Jan Idor Jacobsen and Anders Eckmann deserve special mention.

I would like to take this opportunity to thank Paul Wouters, Programme Leader Virtual Knowledge Studio and Professor of Knowledge Dynamics. I am grateful to you and all the members of the Virtual Knowledge Studio for your hospitality and thoughtful comments regarding my work.

Last but not least, I am greatly indebted to Rasmus. I owe my every achievement to you.

Abstract

The dissertation aims at investigating the changing scholarly communication in general and more specifically the implications of open access on scholarly communication. The overall research question is: What are the effects of open access on scholarly communication? The dissertation consists of five empirical studies of various aspects of the implications of open access on scholarly communication.

The five studies, published as journal articles, are bibliometric studies conducted on three different levels. The first level consists of two studies of a general, more explorative character. The first general study analyses the coverage of open access base resources and the second the use of open access journals in the sciences. The next level of analysis consists of two specific studies that look into two widespread assumptions of the implications of open access. The first is the assumption that the developing countries are great beneficiaries of open access and the second is the belief that open access causes more citations. The third level consists of a concluding, perspectival study. The levels in the thesis to some extent also follow the chronological order of the studies.

Furthermore, the dissertation includes a chapter that presents and discusses the research findings in a theoretical framework. Initially the chapter presents and discusses terminology needed for analysing open access and scholarly communication. Following the necessary definitions and clarifications the theoretical and empirical work of particular interest is presented. The theoretical framework can be divided into two: mappings of scholarly communication and theories of citing. The research findings are summarised in relation to both the overall research question and the theoretical framework. The implications of the research and suggestions for future research are discussed.

Resumé

Denne afhandling undersøger et aspekt af den videnskabelige kommunikation i forandring og fokuserer på implikationerne af, at en voksende andel af den videnskabelige litteratur er frit tilgængelig (også kaldet open access). Det overordnede forskningsspørgsmål er: hvilke effekter er der af denne udvikling for den videnskabelige kommunikation? Afhandlingen består af fem empiriske studier af forskellige aspekter af effekter af open access for videnskabelig kommunikation. De fem studier er alle publiceret i videnskabelige tidsskrifter og er bibliometriske studier foretaget på tre forskellige niveauer. De tre niveauer afspejler desuden kronologien i arbejdet med det overordnede forskningsspørgsmål.

Det første niveau udgøres af to studier, der har generel, mere eksplorativ karakter. Det næste niveau udgøres af to studier af mere specifik karakter, der begge nærmere empirisk belyser to udbredte antagelser om effekten af open access. Den første er antagelsen om, at en af de store effekter af open access er udviklingslandenes større mulighed for synlighed i de videnskabelige samfund. Den anden er antagelsen om, at publikationer, der er frit tilgængelige, bliver læst mere og dernæst brugt mere (målt på antal citationer) end tilsvarende publikationer, der ikke er frit tilgængelige. Endelig består det tredje niveau af et konkluderende og perspektiverende studie, der undersøger forskningspotentialet for videnskabsstudier med den øgede mængde frit tilgængelig videnskabelig litteratur.

Derudover indeholder afhandlingen et sammenskrivningskapitel, der præsenterer og diskuterer resultaterne præsenteret i de fem artikler i en teoretisk ramme. Indledningsvist præsenteres og diskuteres den terminologi, der er nødvendig for at analysere open access og videnskabelig kommunikation. Dernæst præsenteres den eksisterende teoretiske og empiriske litteratur af særlig interesse for denne afhandling. Den teoretiske ramme kan opdeles i forskning vedrørende kortlægninger af den videnskabelige kommunikation samt citationsteorier. Resultaterne fra de fem studier resumeres i relation til både det overordnede forskningsspørgsmål og den teoretiske ramme. Implikationerne af afhandlingen og forslag til fremtidig forskning diskuteres afslutningsvis.

Table of contents

The attached articles

1. Frandsen, T.F. & Nicolaisen, J. (2008). Intra-disciplinary differences in database coverage and the consequences for bibliometric research. Journal of the American Society for Information Science and Technology, 59(10): 1570-1581. Available: http://www.hprints.org/hprints-00326292/en/

2. Frandsen, T. F. (2009). The integration of open access journals in the scholarly communication system: Three science fields. Forthcoming in Information Processing & Management, 45(1), 131-141. Available: http://www.hprints.org/hprints-00326285/en/

3. Frandsen, T. F. (2009). Attracted to open access journals: A bibliometric author analysis in the field of biology. Journal of Documentation, 65(1), 58-82. Available in preprint: http://www.hprints.org/hprints-00328270/en/

4. Frandsen, T. F. (2009). The effects of open access on un-published documents: A case study of economics working papers. Journal of Informetrics, 3(2), 124-133. Available in preprint: http://www.hprints.org/hprints-00352359/en/

5. Frandsen, T. F. & Wouters, P. (2009). Turning working papers into articles: An exercise in micro-bibliometrics. Journal of the American Society for Information Science and Technology, 60(4), 728-739.

The authors of co-authored articles have contributed equally.

List of tables

List of figures

1. Introduction

The purpose of the present chapter is to lay a theoretical framework in which the implications of open access can be analysed. The five attached articles are all empirically based and this chapter enables an analysis of the results from a theoretical perspective.

Consequently, this chapter presents and discusses the overall research question, objective, motivation, data and methods of the dissertation. Necessary definitions are provided before turning to the theoretical framework in which the results of the empirical articles will be analysed.

1.1. Changes in scholarly communication

The conduct of science and scholarship is changing. Bohlin (2004) argues that "the networks of computers, which, over the last decades, have gradually spread across large parts of the globe, have been put to a variety of uses by academic communities" (Bohlin, 2004: 369). He also argues that the introduction of networked computers triggered the radical transformation. Hurd (2005: 7) argues that:

> "[The Internet and the World Wide Web] were products of scientists seeking to find better ways to facilitate sharing of research results; both developments have had a much broader and disruptive impact on the entire system of scientific communication".

The following is a short overview of how these changes to scholarship can be described as it has been done using a wide variety of terms. "Big science" as coined by Weinberg (1961) describes the large-scale enterprise of science in the 20th century whereas "little science" is the individual or small group efforts. According to Borgman (2007: 28) the term "big science" has two different meanings as De Solla Price (1963) shortly after formulated an alternative conceptualization of "big science" as a description of mature

fields with formalised research methods rather than of the size of budgets, staff, machines or laboratories. According to Jankowski (in press) the use of the conceptualisation "cyberscience" goes back as far as 1996, however, the use of it is relatively limited. The concept can be seen as synonomous with "e-science" (e.g. Hackett et al., 2008) which on the other hand is a European version of the American term "cyberinfrastructure" (Jankowski, in press). According to Wouters et al. (2008; 325) "the concept of e-science stresses computational research, processing of huge datasets, video-conferencing, and collaborative research relying on digital communication channels". The term "cyberinfrastructure" was coined by the United States National Science Foundation (NSF) and can be defined as "[i]nfrastructure based on distributed computer, information, and communications technology" (National Science Board, 2008). Consequently, the former term emphasises the practice of science whereas the latter places emphasis on the infrastructure as argued by Wouters et al. (2008).

Finally, "e-research" is a conceptualisation that can be seen to extend the concept of "e-science" to encompass the social sciences and humanities. The term places less emphasis on the size of e.g. machines or laboratories but more on the use of new media, and international and / or multi-disciplinary collaboration. Wouters et al. (2008: 332) argues:

> [E]-science not only claims to be interdisciplinary but also promises to transform the core idea of disciplinarity. While the idea that e-science exist "beyond disciplines" might characterize the discourse about e-science better than its practice, transdisciplinarity is a strong element of e-research."

According to Jankowski (in press) "e-research" can be seen as a successor to the earlier notion of "cyberscience" that is more amenable to the conceptualization of scholarship in the social sciences and humanities.

Borgman (2007: 20) argues that the "e-" prefix as well as the "cyber-" prefix is not used exclusively with this meaning thus the proliferation of terms have not provided an unambiguous conceptualisation of the scholarship in the digital age as she phrases it. Consequently, she prefers the term "information infrastructure".

> Information infrastructure is used here as a collective term for the technical, social, and political framework that encompasses people, technology, tools, and services used to facilitate the distributed, collaborative use of content over time and distance Borgman (2007: 19).

These conceptualizations of the changing conduct of science all try to encapsulate the impacts of information and communication technologies (ICTs) on scholarship. According to Borgman (2007) the impacts are in terms of content and connectivity. An example of developments in ICTs leading to an impact in content and connectivity is electronic publishing. Electronic publishing is a part of the scholarly communication. Electronic publishing is according to Ross and Ross (2002: 480) "a general term embracing all forms of computerized publication, particularly those that deliver text or other materials directly to the consumer's computer screen".

Wang (2003: 1012) overviews the history of electronic publishing and argues:

> In the mid-1980s, a wide variety of electronic publishing and networks mushroomed due to the emerging capability and availability of personal computers, artificial intelligence, software packages, and other innovations.

The advent of the Internet made it possible for research to be shared in entirely new ways. Joint with the technological developments enabling networked dissemination of scholarly communication the development in journal prices led to the so-called "serials crisis" (Albert, 2006). The serials crisis was triggered by the journal subscription fees that have been rising at a rate far above inflation over the past several decades (McCabe 2002; Ortelbach, 2008). Consequently, the rising journal prices led to less access as libraries had to cancel significant journal titles at the same point in time as new technology make it possible to disseminate a scholarly publication by making it instantly accessible anywhere in the world (where there are computers and internet connections).

According to the timeline of what was later to be called the open access movement by Peter Suber (http://www.earlham.edu/~peters/fos/timeline.htm) the first free, online, peer reviewed journals appeared in the late 1980s. Since then the amount of scholarly publications freely available on the Internet has grown considerably (see e.g. Björk,

Roos & Lauri, 2009; Hajjem, Harnad and Gingras, 2005). Both in the form of open access journals as well as openly available versions of articles published in traditional toll-access journals.[1]

Consequently, open access is an element in the changes to scholarship. However, the following section clarifies that apart from being a subject of research into scholarly communication open access is an ideological cause as well.

1.2. Open access as a cause

Open access is both the subject of research into scholarly communication as well as a cause. These two fundamentally different approaches to discussing open access tend to collide. An example of that are the rapid responses on BMJ.com to the article by Davis et al. (2008). Davis et al. (2008) measure the effect of free access to the scientific literature on article downloads and citations. Their results are rather controversial as they find that open access publishing may reach more readers than subscription access publishing. However, they find no evidence of a citation advantage for open access articles in the first year after publication. BMJ allows readers to comment on the article and it attracted quite a lot of responses (more than 15 in about a month). Each respondent is required to declare if they have any competing interests. Under the title "Davis et al's 1-year Study of Self-Selection Bias: No Self-Archiving Control, No OA Effect, No Conclusion" Stevan Harnad (2008a) lists 18 critical comments in relation to the article and declares no competing interests; however, the following day he wishes to correct his declaration of competing interests to the following: "OA Advocate and co-author of several articles reporting OA citation advantage" (Harnad, 2008b). Consequently, he sees himself as both an OA advocate and a researcher investigating the phenomenon. However, it may be easier said than done to fulfil both roles simultaneously.

[1] Although overall OA coverage can be calculated great differences across disciplines should be noted. Examples of remarkably high coverage: A few areas of physics are characterised by almost 100 per cent (Shadbolt et al., 2006) and the CERN institutional repository contains just over half of all documents written since the foundation of CERN in 1954 (Yeomans, 2006).

Partly due to the somewhat biased researchers working within the field and partly due to the very difficult nature of studies of OA numerous tales of OA exist. The effect of OA is very difficult to assess and the great challenge is the determination of causation. A study that illustrates association is much simpler because in order to determine causation a study must control for the effect of other variables. This is a well-known problem in the social sciences. Furthermore, determining the status of a publication as being OA or non-OA is not a trivial task. A publication can be made OA in numerous ways and the publication can exist in an earlier version maybe even with a different title. Finally, there are issues of field specific variations which complicate studies across fields. Adding all these time consuming challenges it is tedious work to determine causation of a potentially very small effect. The perfect study of OA is probably not feasible within a realistic time frame and the realistic methods for OA studies are not unproblematic. Consequently, the ongoing debate on OA is also marked by the fact that some of the participants consider themselves OA advocates.

In the following two tales of OA will be described. Numerous tales exist but these two are exemplary. The two tales will be presented here but not thoroughly discussed as that debate can be found in the attached articles.

The first example is the confidence in OA being of particular value to the developing countries. There are differences in the impact of ICTs on scholarly communication depending on the access provided currently through subscriptions. Scholars from developing countries have limited access to research publications due to expensive subscription costs and the open access movement is challenging the constraint to access. Consequently, researchers in developing countries are often mentioned as major recipients of the benefits when advocating OA (Chan & Kirsop, 2001; Kirsop & Chan, 2005).

Chan and Kirsop (2001) list several advantages for developing countries: (a) free access to research information from developed countries; (b) researchers in developing countries can archive their own research and thus make it available to researchers in developed countries; (c) local research can be distributed. Developing countries are often mentioned when describing the advantages of open access: “One of the great beneficiaries of open access may be users in developing countries, where there are currently some universities with no journal subscriptions at all”

(Wikipedia.org/wiki/open_access, accessed April 2009). Along the same lines Xuemao and Chang (2006) state that "OA will benefit the developing countries in the two-way scholarly communication and accelerate the development of science accordingly." Furthermore, Nicholas, Huntingdon and Jamali (2007: 877) makes the prediction about a journal (Nucleic Acids Research) moving to an OA model "(that) it would be assumed that there is some further growth in the pipeline, predominantly from second and third world users".

Chan (2008) even argues than although authors from developing countries may not publish more in OA journals or cite them more OA journals should be ensured.

> If there should be a study showing that people in developing countries prefer imported bottled water over local drinking water, should efforts to ensure clean water supply locally be questioned?

Consequently, OA is perceived as the best possible solution for the researchers in the developing countries.

The second example is the belief that OA publications attract more citations. A perceived major benefit for authors making their publications available OA is the higher number of citations to OA publications presumably due to higher visibility and /or access. The open access advantage needs an initial clarification of the relations to other measures of usage. The supposed increase in the number of citations received by open access publications compared to (similar) non-OA publications must be kept separated from other measures of usage. Open access would, all other things equal, have a greater potential audience. It would at least be difficult to imagine a smaller audience for a specific journal article if self-archived in an open access repository. An example of other measures of usage is number of downloads (Bollen and Van de Sompel, 2008). The number of citations and the number of downloads are not necessarily causally related meaning that increased potential audience and thus readership does not necessarily imply a greater number of citations. However, there may be relations between the measures in general (e.g. Perneger, 2004; Brody, Harnad & Carr, 2006). Davis et al. (2008) isolates the effects of OA for OA-articles and non-OA articles and finds that OA-articles are more downloaded than non-OA articles in the same journals.

They are, however, not more cited (although a study with a longer citation window would have been preferable).

The OA citation advantage is investigated by Lawrence (2001) who analysed the effects of OA on citation impact for a sample of conference documents and found freely available papers to have greater impact. However, his study was restricted only to computer science which makes it difficult to generalise. This effect, "the open access citation advantage" has since been confirmed on larger samples by Antelman (2004), Harnad and Brody (2004), Hajjem et al. (2005), Hajjem, Harnad and Gingras (2005), Metcalfe (2005), Henneken et al. (2006). However, one should be careful making causal arguments as pointed out by e.g. Craig et al. (2007) stressing that in order to conclude that OA publication causes more citations, we need data to illustrate causation, not just association. The increased number of citations could be caused by other factors than the increased visibility of the work as stressed in the work by Kurtz et al. (2005), Davis and Fromerth (2006), Moed (2007), Davis (2008), Davis et al. (2008), Gaule and Maystre (2008) and Norris, Oppenheim and Fytton (2008). The aforementioned studies all point to other possible explanations of the OA advantage (e.g. 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). Regardless of these studies the Wikipedia article on OA claims the following:

> The main reason authors make their articles openly accessible is to maximize their research impact. A study in 2001 first reported an Open Access citation impact advantage, and a growing number of studies have confirmed, with varying degrees of methodological rigor, that an open access article is more likely to be used and cited than one behind subscription barriers. For example, a 2006 study in PLoS Biology found that articles published as immediate open access in PNAS were three times more likely to be cited than non-open access papers, and were also cited more than PNAS articles that were only self-archived. Recently, this result has been challenged as possibly due to a quality bias (Accessed April 2009)

The OA advantage is presented as an empirical fact and only one reference is mentioned to challenge that empirical fact. The study by Gaule and Maystre (2008) is mentioned, however, the exemplarily thorough study by Moed (2007) is not even mentioned in the

reference list of the article. One of the major contributors to the article is Stevan Harnad who is listed with no less than 66 contributions to the article. He has knowledge of the studies challenging the OA advantage as an empirical fact (e.g. see his blog from September 7, 2007: http://openaccess.eprints.org/) and could contribute to representing the studies of OA advantage more comprehensively.

Summing up, various aspects of OA have been investigated in the existing literature, however; the existing literature on OA is to some extent affected by the strong opinions regarding OA. Related to the strong opinions is the understanding of the OA community as a movement. The OA community does not only share the professional interest in OA but also the same goals of achieving more OA.

1.3. Motivation and objective

The present dissertation is motivated by an interest in the changes to scholarship and what role OA plays in these changes. Open access is definitely changing the understanding of the term publication. Scholarly publications can be made immediately public on the Internet, however, to what extent that changes the scholarly communication is another question.

Researchers working within the field of OA struggle with the difficult nature of the studies. Causation is extremely difficult to determine and association is of very little use. On top of that the existing literature on OA is characterised by a number of researchers trying to promote the cause of OA.

The objective of the dissertation is to provide empirical, bibliometric analyses that investigate some of the existing assumptions regarding OA. Hopefully, the results of these analyses can contribute to providing the grounds for a discussion of the advantages and disadvantages of OA as nuanced as possible.

1.4. Research questions

The motivation and objective lead to the following overall research question:

What are the effects of open access on scholarly communication?

This research question can be investigated on several levels and the dissertation consists of three levels of analyses. The first level consists of two studies of a general, more explorative character. The next level of analysis consists of two specific studies that both point back to the aforementioned tales of OA. The third level consists of a concluding, perspectival study. The levels in the thesis to some extent also follow the chronological order of the studies.

The main research question can be elaborated into sub-questions. The general studies focus on introductory studies of open access as an element in scholarly communication.

1. The first study deals with open access as an alternative database and to what extent open access based resources as a database reproduces the existing databases that typically have available and consistent indexing policies. The study treats the following question:

   How does OA affect the visibility of research?

2. The second study focuses on the assumption that open access is a particular science phenomenon. The greatest number of OA journals and archives are found within the sciences that are typically viewed as leaders within open access. However, few studies exist on the differences within the sciences. The second study deals with the following question:

   How can the integration of open access in the existing publishing system be described in the sciences that are the leaders in establishing OA journals?

Two of the studies in the thesis are specific studies focusing on two of the most well-known tales of OA.

3. The third study looks into the assumption that OA is of greatest benefit to the developing countries. The study is a supplement to existing author analyses based on interviews and questionnaires and the main research question of the study is:

   What are the implications of OA for the developing countries in terms of using OA available research and publishing their own research in OA journals?

4. The fourth study deals with the most controversial OA tale: the open access citation advantage. A number of studies exists that focus on the potential OA advantage of journal articles. However this study uses a different focal point:

   What are the implications of open access for publication types that have received little attention before the Internet?

Finally, the thesis includes a study of more perspectival character which tries to open up new perspectives within the field of OA.

5. What can open access provide of new possibilities in terms of research into scholarly communication?

1.5. Data and methods

The focus of this project is the changing scholarly communication. Information science offers interesting perspectives on how to analyse the impact of ICTs on scholarly communication. According to Wouters et al. (2008: 340) information science can offer two approaches to studying the use of ICT in scientific communities.

> In Information science, we can differentiate between "analyses" of the "impact" of ICT on traditional scholarly practices (collaboration, publishing behavior [Lawrence 2001, Wouters & de Vries 2004], the emergence of new

> scholarly practices (email, chat, on-line peer review), and new ways of studying scholarly practices (both using Web data [hyperlinks]] as well as digitized bibliometric data [Chen & Lobo 2006]).

Following the distinction by Wouters et al. (2008), all five sub-questions relate to traditional scholarly practice although sub-question 5 can also be placed in the third approach as it relates to new ways of studying scholarly practices by using digitized bibliometric data. All approaches mentioned by Wouters et al. may include bibliometric data. According to Borgman and Furner (2002: 4) bibliometrics is particularly suitable for studying scholarly communication:

> Bibliometrics offers a powerful set of methods and measures for studying the structure and process of scholarly communication. Citation analysis, the best known of bibliometric approaches, has become more sophisticated, and the advent of networked information technologies has led to quantitative and qualitative advances in other bibliometric methods.

Bibliometrics is defined by Tague-Sutcliffe (1992: 1) as "the study of the quantitative aspects of the production, dissemination and use of recorded information". Bibliometrics is strongly related to informetrics and scientometrics which Tague-Sutcliffe (1992: 1) also provides definitions for. "Informetrics is the study of the quantitative aspects of information in any form, not just records or bibliographies, and in any social group, not just scientists." "Scientometrics is the study of the quantitative aspects of science as a discipline or economic activity." Consequently, scientometrics would be the most appropriate term in this context, however; the terms are often used as synonyms and being the oldest of the terms, bibliometrics is often used where scientometrics would be more precise. Bibliometrics is the term used here.

The short version of what bibliometricans do is provided by Cronin (2001: 1) stating that bibliometricians count and measure things. More specifically:

> In bibliometrics, as informetrics and scientometrics, the derived measures or metrics are typically counts of the frequencies with which events of specified types are observed to occur, which (once expressed as ratios of the

> total number of observed events) may be considered as probabilities of occurrence. (Borgman & furner, 2002: 7).

According to Cronin (2001), the efforts have traditionally been concentrated on tracking highly visible and objective indicators of scholarly activity. Typically the focus has been on publications and citations. However, a number of inputs and outputs can be used for bibliometric studies and he provides the following examples: journals, acknowledgements, scientific manpower, federal funding patterns and rates of patenting.

1.6. The structure of the dissertation

Section 2 is occupied primarily with definitions as it defines open access and subsequently an open access publication. In order to analyse the effects of open access as means in scholarly communication we need to establish what scholarly communication is and how it can be analysed. Section 3 defines scholarly communication and the theoretical and empirical framework in which changes in scholarly communication can be analysed. Summaries of the articles attached are provided in section 4. Finally, conclusions are being drawn in section 5 on the basis of the articles and the theoretical framework.

2. Open access

The term open access can be seen as synonymous with the term free online scholarship, however, as stated by Suber (2004a):

> I use [free online scholarship] as the generic term for scholarly literature in the sciences or humanities available free of charge on the internet. I turned to a new term only reluctantly because there was no general term already accepted for this kind of literature, although since the launch of the Budapest Open Access Initiative (2/14/02), the term open access has spread widely.

Defining open access is not a trivial task as the demarcation of the term tends to be somewhat unclear. The following describes and discusses definitions.

2.1. The open access movement

The modern open access movement is tied to the changes in academic publishing. The radical transformation of scholarly publishing was set off by the introduction of computerized networks, although, Willinsky (2006: 30) notes that the "open access idea is not just a child of these new publishing technologies" and provides a historical framework for the open access movement. He argues that open access can be seen as:

> [T]he next step in a tradition that includes the printing press and penny cost, public libraries and public schools. It is a tradition bent on increasing the democratic circulation of knowledge, with a lineage that can also be traced back, for example, to the "invisible colleges" of the seventeenth century, which were comprised of informal study clubs that would gather in coffee houses, otherwise known as "penny universities" Willinsky (2006: 30).

The movement consists of various aspects of which Domienguez (2006) mentions the political, technological, legal and economic. For the movement it is central to agitate that peer-reviewed journal literature is made freely available to the public through the internet. Or as they see it: OA is an attempt to combine the willingness of scholars to publish the fruits of their research in scholarly journals without payment, for the sake of inquiry and knowledge and the means of world-wide electronic distribution of the new technology.

The political actions taken to support open access have resulted in a number of declarations. According to Suber (2008) the Budapest (February 2002), Bethesda (June 2003), and Berlin (October 2003) definitions of open access are the most central and influential for the OA movement. They are collectively termed the BBB statement. The Budapest Open Access Initiative consists of a statement of principle, a statement of strategy, and a statement of commitment. According to the website of the initiative hosted by Open Society Institute and the Soros foundations network (http://www.soros.org/openaccess/index.shtml) the declaration was formulated at a meeting organised in Budapest by the Open Society Institute. The initiative was initially signed by the Budapest participants and signatures were invited of all interested people and organisations. Open access is defined by the Budapest Open Access Initiative (2002) as follows:

> There are many degrees and kinds of wider and easier access to this literature. By 'open access' to this literature, we mean its free availability on the public internet, permitting any users to read, download, copy, distribute, print, search, or link to the full texts of these articles, crawl them for indexing, pass them as data to software, or use them for any other lawful purpose, without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. The only constraint on reproduction and distribution, and the only role for copyright in this domain, should be to give authors control over the integrity of their work and the right to be properly acknowledged and cited.

However, although this definition is widely used and cited (e.g. Hall, 2008; Willinsky, 2006) the Bethesda and Berlin statements are more recent. Suber (2003, June) describes how the Bethesda statement was drafted during a one-day meeting held at the

headquarters of the Howard Hughes Medical Institute in Chevy Chase, Maryland. He states that he is not "an official spokesperson for this statement, just a participant in the conference that drafted it" (Suber, 2003, June). The purpose of the statement was to promote transition to open access publishing within the biomedical research community. Attendees signed the statement on the basis of group consensus. Open access is defined by the Bethesda Statement on Open Access Publishing as follows:

> An Open Access Publication [1] is one that meets the following two conditions:
>
> The author(s) and copyright holder(s) grant(s) to all users a free, irrevocable, worldwide, perpetual right of access to, and a license to copy, use, distribute, transmit and display the work publicly and to make and distribute derivative works, in any digital medium for any responsible purpose, subject to proper attribution of authorship [2], as well as the right to make small numbers of printed copies for their personal use.
>
> A complete version of the work and all supplemental materials, including a copy of the permission as stated above, in a suitable standard electronic format is deposited immediately upon initial publication in at least one online repository that is supported by an academic institution, scholarly society, government agency, or other well-established organization that seeks to enable open access, unrestricted distribution, interoperability, and long-term archiving (for the biomedical sciences, PubMed Central is such a repository).
>
> Notes:
>
> 1. Open access is a property of individual works, not necessarily journals or publishers.
>
> 2. Community standards, rather than copyright law, will continue to provide the mechanism for enforcement of proper attribution and responsible use of the published work, as they do now.

Finally, the most recent of the three definitions is the Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities. Following a three-day conference in the Harnack House of the Max Planck Society in Berlin the declaration was signed by the invited attendees from German and international research organisations. Organisations that commit to implementing this definition of open access can sign on to the declaration with the following definition of open access:

> Open access contributions must satisfy two conditions:
>
> The author(s) and right holder(s) of such contributions grant(s) to all users a free, irrevocable, worldwide, right of access to, and a license to copy, use, distribute, transmit and display the work publicly and to make and distribute derivative works, in any digital medium for any responsible purpose, subject to proper attribution of authorship (community standards, will continue to provide the mechanism for enforcement of proper attribution and responsible use of the published work, as they do now), as well as the right to make small numbers of printed copies for their personal use.
>
> 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.

These three definitions of open access differ only slightly and because they agree in the substance they are referred to collectively as the BBB statement (Suber, 2008a). According to Suber (2004b) "[n]early all OA proponents agree on the BBB definition", nevertheless he acknowledges that the term is diluted and "true open access" is debatable. He argues that among other reasons for this dilution is that the definition leaves room for variation. An example of that variation is noted by Harnad (2006):

> My definition is the same as that of the Budapest convention: that open access gives free online full-text access to peer-reviewed literature. This definition is missing two important words though, immediate and permanent.

Consequently, he argues that it is not adequately included in the definition when and how long access should be provided. Another element in the definition that is debatable is whether or not a publication should be peer-reviewed to be included. In the following these and other elements of variation are explored in further detail.

2.2. Open access as principle

As pointed out by Moed (2007: 2048) the term open access has different meanings when it comes to the practical implementation of the definitions above:

> It is used to indicate a particular business model of scientific publishing, in which essentially the authors of articles published in a journal pay the costs of the publication, and their full texts are freely accessible once they are published. But the term "Open Access" is also used to indicate open or free accessibility of scientific documents in general, regardless of whether these are published in a journal running under an Open Access model, or published in a journal applying other business models but also (often after several months) deposited in a freely accessible archive such as a personal website or an institutional depository, or as pre-prints in a freely accessible pre-print server.

A recent example of defining open access as principle is provided by Hall (2008: 3):

> By open access, I mean access that is digital, online, and free of charge to those able to connect to the Internet, without having to pay subscriptions either to publish or to [pay per] view, in its purest form, anyway.

Willinsky (2006: xii) argues that open access is a part of a broader term called the access principle which he defines as:

> A commitment to the value and quality of research carries with it a responsibility to extend the circulation of such work as far as possible and ideally to all who are interested in it and all who might benefit from it.

He argues that open access is not free access and that the open access movement "is not operating in denial of economic realities" (Willinsky, 2006: xii). He further argues that the movement is concerned with increased access and concerned with a long, withstanding scholarly tradition of extending the circulation of knowledge.

In common for the definitions of open access tied to the principle of open access is that they focus on the beneficiaries of open access and less on how the access is to be financed. The lack of financial aspects in the definition leads to difficulties in discussions on open access because one can hardly disagree or be against open access in this sense.

2.3. Open access models

Willinsky (2006: 211) argues that

> [O]nly with a loosely defined approach to open access archiving and publishing can one begin to capture the variety of and variation in the means that are now being used to increase access to scholarship and research.

He groups the variations according to how they are financed and the nature of the access they provide and identifies ten flavors or forms of open access. He points out that some of these flavors contravene some of the existing definitions of OA (e.g., Budapest Open Archives Initiative, 2002; Bethesda Statement 2003), and should be thought of as means of increasing access to research and scholarship. In other words, the ten forms of OA publishing and archiving described by Willinsky provide better

access than traditional scholarly publishing models. The ten forms are the following (reprinted from Willinsky, 2006: 211-216):

Table 1. Ten flavors of open access to journal articles

Type of open access	Economic models	Journal or portal example
Home page	University department maintains home pages for individual faculty members on which they place their papers and make them freely available.[a]	http://www.econ.ucsb.edu/~tedb/
E-print archive	An institution or academic subject area underwrites the hosting and maintenance of repository software, enabling members to self-archive published and unpublished materials.[a]	arXiv.org E-Print Archive
Author fee	Author fees support immediate and complete access to open access journals (or, in some cases, to the individual articles for which fees were paid), with institutional and national memberships available to cover author fees.[a]	BioMed Central
Subsidized	Subsidy from scholarly society, institution and/or government / foundation enables immediate and complete access to open access journal.[a]	First Monday
Dual-mode	Subscriptions are collected for print edition and used to sustain both print	Journal of Postgraduate

	edition and online open access edition.[a]	Medicine
Delayed	Subscription fees are collected for print edition and immediate access to online edition, with open access provided to content after a period of time (e.g., six to twelve months).	New England Journal of Medicine
Partial	Open access is provided to a small selection of articles in each issue—serving as a marketing tool—whereas access to the rest of the issue requires subscription.	Lancet
Per capita	Open access is offered to scholars and students in developing countries as a charitable contribution, with expense limited to registering institutions in an access management system.	HINARI
Indexing	Open access to bibliographic information and abstracts is provided as a government service or, for publishers, a marketing tool, often with links to pay per view for the full text of articles.	ScienceDirect
Cooperative	Member institutions (e.g., libraries, scholarly associations) contribute to support of open access journals and development of publishing resources.[a]	German Academic Publishers

a. Supports ‘‘open access’’ as defined by the Budapest Open Access Initiative (2002) and Bethesda Statement on Open Access Publishing (2003), although some users may impose restrictions that fall outside these definitions (e.g., Bethesda Statement: “Grant(s) to all users a free, irrevocable, worldwide, perpetual right of access to, and a license to copy, use, distribute, transmit and display the work publicly and to make and

distribute derivative works, in any digital medium for any responsible purpose, subject to proper attribution of authorship, as well as the right to make small numbers of printed copies for their personal use" [2003]).

The ten flavours by Willinsky overlap partially with the widely used grouping of open access forms, the so-called green and gold strategies.[2] Both approaches to grouping variations of open access are based on what kind of access is provided and how it is financed. Open access can be seen as consisting of two phases or strategies for achieving open access: open access journals and self-archiving (Brody and Harnad, 2005). Open access journals or 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. Table 2 provide an overview of the OA publishing models. The table is an excerpt from (Dominguez, 2006).

[2] A number of color codes have later been added to the color scheme. See Jenkins et al. (2007) for an overview of an abridged color scheme.

OA publishing models	**Descriptions**	**Examples**
All free OA	No author or reader fee. Online only.	Living Reviews of Solar Physics. Former JHEP
Partial OA	Some articles in an issue are OA	Many publishers use this for promoting a journal to a wider audience
Per Capita country	OA made available to country based on per capita income. Mainly in developing countries.	Oxford University Press, Proceedings of the National Academy of Sciences of the USA
Paid OA	Authors pay to publish. Readers have free access to these works.	Springer Open Choice, Public Library of Sciences, American Institute of Physics, Institute of Physics

Table 2. Open access journals models: the gold way

Self-archiving or the green self archiving strategy comprises of persuading authors to self-archive the articles they publish in traditional toll-access journals in open access archives. The many archives are to be searched collectively provided that they comply with a standard. The green way is free for both authors and readers. In the EPIC (2004) a distinction is made between archives and repositories which refers to an archive as a collection of material published in journals and a repository as a collection of material including both un-published and published material. However, this distinction is often not used and the terms are seen as synonyms. According to Dewatripont et al. (2006) two types of archives and / or repositories exist. An overview is provided in table 3.

	Descriptions	**Examples**
Subject-based archives	Collecting and providing access to articles and documents in a specific discipline. They have emerged in domains with long-standing tradition for exchanging pre-prints and where speedy publication is imperative.	High-energy physics and mathematics (arXiv), computer science (NCSTRL).
Institutional repositories	Preserving, disseminating and managing the scientific institution, typically including theses and dissertations, working papers, conference papers and published articles.	CERN Document Server

Table 3. Open access archives models: the green way.

The grouping of open access in open access journals and self-archiving with a number of variations under each heading is less inclusive than the ten forms identified by Willinsky. Some of the additional forms identified by Willinsky can be added under one of the two headings (e.g. dual mode or home page) and the additional forms that cannot be placed under a heading (e.g. indexing and cooperative) describe open access but not to scholarly publications and is thus of less interest in relations to existing definitions of open access. Consequently, the terminology used in these two approaches to group open access is not in conflict with each other and they both offer relatively precise and operational definitions of the forms of open access.

2.4. Defining a scholarly publication

First of all, it is necessary to determine which publications can be embraced by the open access term. According to Oxford English Dictionary (online edition) a publication is "[a] published work; a book, newspaper, etc., produced and issued for public sale or distribution; a text made publicly accessible or available in electronic form" (accessed April, 2009). Consequently, the key issue is that the document is made public. The Budapest, Bethesda, and Berlin definitions of open access include to a varying degree a characterization of what is provided access to. The Budapest definition states the following regarding what publications an author could make OA:

> Primarily, this category encompasses their peer-reviewed journal articles, but it also includes any unreviewed preprints that they might wish to put online for comment or to alert colleagues to important research findings.

The Bethesda definition merely states that "Open access is a property of individual works, not necessarily journals or publishers." The Berlin definition is slightly more specific by outlining examples of open access contributions:

> Open access contributions include original scientific research results, raw data and metadata, source materials, digital representations of pictorial and graphical materials and scholarly multimedia material.

Summing up the BBB definitions of a contribution that can be open access we can say that a contribution is not necessarily peer reviewed and is typically a journal article but can also be e.g. data. However, the BBB definitions do not provide a precise definition.

Borgman (2007) refers to a leaflet with a much narrower definition of a publication which outlines open access as peer reviewed journal articles, technical reports, theses and working papers. However, the problem of restricting the definition to peer reviewed publications is that it can be exceedingly difficult to determine the character and quality of the peer review process. Primarily, because an assessment of a peer review process needs to include the "luck of the reviewer draw" (Cole & Cole, 1981). The specific reviewers heavily influence the outcome of the process and assigning other

reviewers to the same publication can easily result in a different outcome. Consequently, a peer reviewed publication is not necessarily of better quality than an un-refereed publication. Numerous examples exist that questions the peer reviewed publication being of better quality than an un-refereed publication. The Nobel Award winner, Albert Einstein, is a well-known example of a scientist struggling with peer review (e.g. Pais, 1982). Another example is provided by Gans and Shepard (1994) finding that papers in economics have been rejected that later earned the author the Nobel award. Even in a perfect world with flawless peer reviews, un-refereed literature can fill a vital function. Long publication lags can have the consequence that an un-refereed version of the publication circulates for several years before the official peer reviewed version is published. Most journals have decreased the lag-time from submission to appearance considerably by handling all submitted publications online including the peer review process and checking of author proofs (Jones, 2003). Ellison (2002a) and Ellison (2002b) show that economics as well as other disciplines have experienced a slowdown in submit to accept times over the last three decades. The major cause is that authors are required to revise their papers more times and more extensively than in the past. Azar (2005) reports increases in the first response time of economics journals over the last four decades and he even argues that the efficient circulation of un-refereed versions may contribute to the longer lags because quick publication becomes less important as the research is distributed prior to publication, the optimal number of revisions increases. Consequently, cause and effect can be difficult to determine, however, it remains that a peer reviewed publication is not necessarily of better quality than an un-refereed publication.

Borgman (2007: 98-99) discusses what constitutes a publication and argues that a number of issues needs to be addressed in terms of prior publications. Kling (2004) provides a precise definition of the terminology regarding working papers, pre-prints and e-prints), new genres (data repositories, electronic books and theses including features not replicable in print form) and informal scholarly communication genres (e.g. blogs, discussion links and RSS feeds). However, it is practically impossible to define precise criteria for determining what constitutes a publication. It is relatively straightforward to give an unambiguous example of a publication (e.g. a peer reviewed journal article) and likewise of a document not qualifying as publication (e.g. the personal notes of a scholar). But between these extremities it is impossible to draw definite lines and consequently, a vague definition of a publication is inevitable.

Secondly, another issue to resolve is what constitutes "scholarly" or "scientific". Little clarification is offered in the existing definitions of open access. According to Oxford English Dictionary (online edition) scholarly is defined as "[p]ertaining to, or characterizing, a scholar; befitting, or natural to, a scholar; learned, erudite" (accessed April, 2009). Scientific is defined as: "Of persons, books, institutions, etc.: Occupied in or concerned with science or the sciences" (accessed April, 2009). In immediate continuation of that definition is the definition used by Borgman (2007: 47) who argues that "[s]cholarship is an inherently social activity, involving a range of public and private interactions within a research community". Consequently, a scholar can be defined as a member of a research community and does thus not have to be affiliated with a university. The important issue is whether he or she is occupied with doing research.

The demarcation of a scholarly publication should be made in relation to a given purpose. For research evaluation purposes one definition may be appropriate and for the purpose of research dissemination another definition of a scholarly publication may be of more value. However, for the purpose of this dissertation, a scholarly / scientific publication is a document made public containing research. This definition implies that unpublished documents can qualify as publications if they are made public (e.g. on the internet) and the author is a member of a research community. However, it should be noted that this definition excludes informal as well as formal communication if it is not containing research. Consequently, an example of a document type that would often be excluded using this definition is a blog. Blogs may contain research but often it is "science news straight from the source" (www.scienceblog.com).

2.5. Defining access

Culnan (1985) argues that accessibility is a multidimensional concept encompassing:

1. Physical access to the source
2. The interface to the source
3. The ability to physically retrieve potentially relevant information

The second and third dimension is usually not included in a definition of open access, however, nuances have been added to the first dimension of the concept of access as described by Culnan. Probably due to the fundamentally different situation in 1985, Culnan does not elaborate on the concept of physical access. However, as OA involves providing online access physical access can imply reading but it can also imply more than just reading which pertains to copyright and patent issues.

Suber (2003) introduced a terminology for the two component parts of open access: (1) the kind which removes price barriers alone and (2) the kind which removes price barriers and at least some permission barriers. He called them "removal of price barriers" and the "removal of permission barriers".

A new terminology was introduced in the spring of 2008 by Stevan Harnad and Peter Suber. The suggested terms were: "weak OA" and "strong OA" for these two species. However, due to the negative character of the so far suggested terms a new set of terms has been introduced namely "gratis" and "libre". As Suber (2008) argues: "They are accurate, neutral, and descriptive. In the neighboring domain of free and open source software, they exactly express the distinction I have in mind".

According to Harnad (2008c) the gratis/libre distinction is of great importance "because it is critical to the strategy for successfully achieving OA (of either kind) at all. There is still very little OA today, but most of what OA there is is gratis, not libre". However, for the purposes of this dissertation the distinction is not a central issue as the focus is primarily on the first component part (gratis) and not on the second (libre).

3. Scholarly communication

OA can be seen as an element in the changes of the conduct of science and to scholarship in general. In order to analyse the effects of OA as means for scholarly communication we first need to establish what scholarly communication is. Scholarly communication is shaped by the characteristics of the organisation of research. Whitley (1991, 2000) argues that the sciences[3] are distinguished from other sorts of work organisations by a number of particular features that provide the unique characters of science. First of all the output produced by scientists are primarily consumed by people within the discipline e.g. peers. Thus standards and objectives are primarily set internally. Science is characterised by primarily legitimising results by publishing which means that the results of a researcher is not recognised until they are published. Science is furthermore characterised by a high degree of mutual interdependence among scientists. Scientists produce intellectual novelties which have to be evaluated against common background assumptions and established knowledge. The knowledge produced refers to the existing knowledge the work is built upon. The significance of the reciprocal interdependence between researchers means that communication systems are crucial. The formal communication media in science are reinforced by the legitimating role as these media control knowledge claims and facilitate the flow of competently produced results. The formal communication system is central in gaining reputations and hence accessing rewards. The legitimising and control functions of the publication system in science confirm that analyses of the organisation and operation of the communication system is central in analysing the communication in science.

Consequently, in order to analyse changes in scholarly communication we need to define scholarly communication and through that find means to do analyses within that field. However, placing the attached articles within a theoretical framework is not necessarily as straightforward as it may appear because it cannot be done without considering fundamentally different perspectives on science and science studies.

[3] Whitley (2000: 7): "[t]he term "science" is used here in a broad sense to refer to all forms of modern scholarship, rather than just to the natural sciences".

3.1. Defining scholarly communication

Whitley argues that communication is a key element of research and the importance of communication has been pointed out by several other researchers (e.g. Meadows, 1998; Paisley, 1984). Garvey (1979: 9) stresses that "communication is the essence of science" and "science is […] a social system of which interactive communication is the salient feature". Griffith (1990: 31) argues that scholarly communication is interesting as means to investigate underlying general social and cognitive processes in science. Cronin (2003: 1) states that:

> Knowing how scientists work, how they interact with their peers and publics, is not just intrinsically interesting to ethnographers, sociologists of science, and sundry others, but has a bearing on the development of effective academic information resources and information support systems.

The characteristics of the sciences identified by Whitley from an organisational perspective are in line with the majority of existing definitions of scholarly communication that all focus on communication among scientists. However, there are other approaches. Scholarly communication does not have to be defined as restricted to communication among peers. Jankowsky (2009: 7) is an example:

> [S]cholarly communication is viewed as the presentation of research findings to an audience external to the research project, home department or institution of the researcher, for the purpose of sharing and contributing to knowledge.

For existing literature dealing with scholarly communication beyond communication among scientists the reader is referred to Lewison, et al. (2008) and Lewison and Sullivan (2008) for examples of studies focusing on how medical research is mentioned in the media and incorporated in clinical guidelines. The reader is also referred to Lucas and Willinsky (in press) or Zuccala (2009) for recent examples including laypersons.

Garvey (1979: 9) describes how he and his colleagues intended to analyse the flow of scientific information from a standpoint of "science conceived as a social system of

which communication is the salient feature". Garvey (1979: 9) defines scientific communication as:

> "[T]he scientific information exchange which takes place among scientists, not information commerce between science and technology or between science and society in general, even though the latter are aspects of communication in which scientists sometimes find themselves involved" (Garvey, 1979, 9).

Borgman (2007: 48) relies heavily on the definition also used by Garvey, but stresses that one should be careful conflating scholarly communication and scholarly publishing. She defines scholarly communication as:

> The term scholarly communication is used here in the broader sense to include the formal and informal activities associated with the use and dissemination of information through public and private channels.

3.2. Studying scholarly communication

Scholarly communication can be analysed using a number of different approaches and Cronin (2003: 1) provides a number of examples:

> Structural-functionalist accounts of how science works as a social system (Merton, 1976) coexist (and sometimes compete) with thickly descriptive laboratory accounts of science-in-action (e.g., Latour & Woolgar, 1979), discipline-specific mappings of communicative processes (e.g., Garvey & Griffith, 1971) and painstaking socio-historical accounts of scientific practice (e.g., Shapin, 1994).

This section presents and discusses various models of understanding scholarly communication as a field of research. The various models enable laying a theoretical framework for the dissertation. The first model is based on a distinction between process and structure. The second is based on a distinction between sociology of

science and sociology of knowledge. Thirdly, a model of four research traditions is presented and finally, a schema for studying social and cognitive processes in science is presented.

3.2.1. Process and structure

Lievrouw (1990) distinguishes between studies of structure and process. She defines structure as networks of relationships among scholars in a field, such as those revealed by sociometric or bibliometric analyses. Lievrouw (1990) questions why processes are being investigated through analyses of structures. Lievrouw (1990: 59) uses the invisible college as "typical of constructs that describe processes yet are founded on the study of structures".

Following Smelser (1988) analyses of structure are dedicated to discovering non-random regularities and systematic relationships. Smelser (1988: 103) even argues that "the idea of social structure is at the very heart of sociology as a scientific enterprise". He provides numerous examples of social structures: Some macroscopic conceptualisations of structure (structure as arising from considerations of societal survival or the effective functioning of society; structure as arising from domination and coercion) and some microscopic conceptualisations of structure. Consequently, the distinction between structure and process is based on whether or not a study can discover non-random regularities and systematic relationships. Quantitative methods and qualitative methods are two main types of sociological research methods. Sociologists use quantitative methods such as social statistics or network analysis to investigate the structure of a social process or describe patterns in social relationships. Sociologists use qualitative methods such as focused interviews, group discussions and ethnographic methods to investigate social processes. Consequently, studies of structure and process can be characterised in terms of the degree of ability to discover non-random regularities and systematic relationships, however, both types of analyses aim to investigate a social process. A study of structure without including the social process is thus merely a demonstration of a technique.

Lievrouw (1990) provides the use of clustering techniques to construct bibliometric maps as an example of studies of structures and argues that few research questions may

be answered by studies of structures alone. Her answer is to carry out more in depth work in relation to these studies.

Borgman (2007) also distinguishes between process and structure; however, she defines the terms slightly differently. She argues that "[a]nother sense of structure is implied when the term scholarly communication is used to encompass research universities, libraries and publishing" (Borgman, 2000: 414). Process refers to the use and dissemination of scholarly knowledge and structure refers to the knowledge constructed through such use and dissemination. Borgman argues that due to the different variables, levels of analysis, theories and methods studies on process and structure are difficult to compare. According to Borgman (2007: 56-57) research on processes is concentrated within social studies of science with some work in information studies whereas research on structures is scattered among information studies, bibliometrics, webmetrics with some work in social studies of science. Borgman and Furner (2002: 5) situate bibliometrics within studies of scholarly communication:

> Bibliometrics is now an accepted method in sociology of science (J. R. Cole, 2000; Cronin & Atkins, 2000; Merton, 2000), especially by scholars whose inquiries are well served by quantitative methods and structural approaches. Others prefer more qualitative methods and more interpretive or constructivist approaches to the study of scholarly communication.

Park (2008) shows in figure 1 how research on structures as well as processes can be related to science and technology studies (STS). The terms, structure and process, are defined drawing on both Borgman (1990), Borgman (2000), Borgman and Furner (2000), and Lieuvrouw (1990). Park argues that there is an intersection of scholarly communication processes and sociology of scientific knowledge. Furthermore, there is an intersection of scholarly communication structures and sociology of science.

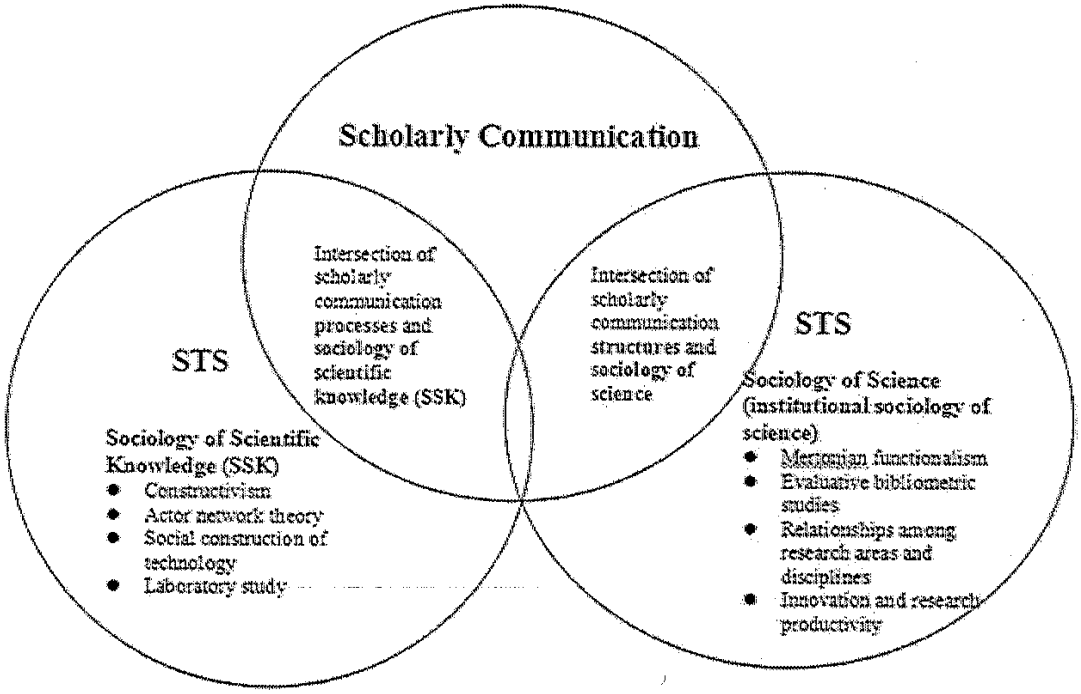


Figure 1. Map of domain intersection between scholarly communication and STS (Park, 2008).

Consequently, the division between sociology of knowledge and sociology of science lies underneath the somewhat problematic distinction between process and structure. Finally, Barjak (2006: 1351) argues that the two most important models of scientific knowledge creation developed in science studies are the sociology of science and sociology of knowledge.

3.2.2. Sociology of science and sociology of knowledge

Sociology of science or as Hess (1997) argues as a more precise term: the institutional sociology of science has its roots in the sociology of occupations and science is seen as a form of occupation. Ben-David and Sullivan (1975: 203) argue that "sociologists of science have concentrated on this characteristic of science as a tradition and as an institution".

No single theoretical or methodological assumption dominates the sociology of science and consequently the field can be difficult to describe it as a whole. Jasanoff et al. (1995: 25) describes it as "the rich terrain of science studies". According to Zuckerman (1988: 513) the research agenda of sociology of science falls in three categories:

> [F]irst, problems concerning the social and cognitive organization of scientific work; second, problems in the sociology of scientific knowledge; and third, the reciprocal connections between science and its social contexts – the social consequences that flow from science and the influence of other social institutions and culture upon science.

Zuckerman (1988: 513) argues that the theoretical differences in the sociology of science reflect differences in the conceptions of science. Callon (1995) identifies four models for the dynamics of science: (a) science as rational knowledge, (b) science as competition, (c) science as sociocultural practice, and (d) science as extended translation.

Sociology of scientific knowledge was developed in opposition to philosophy of science and sociology of science. Zuckerman (1988: 541) argues that sociology of science developed in to two streams:

> [T]hose emphasizing social influences on the structure and development of scientific knowledge and those focusing on the social construction of knowledge itself. The former can be described as "structural" and the latter "constructionist" studies, these covering a great variety of inquiries, including analyses of employing relativist and constructionist perspectives, those treating the impact of social and professional "interests," and discourse and text analysis.

Consequently, sociology of science and sociology of knowledge entails a number of fundamentally very different approaches and the distinction can be difficult to employ as theoretical framework for open access studies. Barjak (2006) uses the distinction as analytical framework for studying informal communication, however; he merges it with the arrays of studies identified by Lievrouw (1988).

3.2.3. Four research traditions

An alternative model for distinguishing approaches to scholarly communication research is the four research traditions identified by Lievrouw (1988). They can be characterised according to the assumptions made about scientific information and the following table is from Lievrouw (1988).

Category	Information assumptions	Specialty	Methods
Artifact studies	Information as commodity: value independent of use	Communication, information science	Bibliometrics
User studies	Information as commodity: value depends on user needs	Information science, communication	Bibliometrics, user surveys
Network studies	Information as social link: value in coherence of social networks	Communication, sociology of science	Network analysis: sociometrics and surveys
"Lab" science	Information as construction: value in the meanings and perceptions of individuals	Sociology of science, sociology of knowledge	Fieldwork: interviews, participant observation

Table 4. Programs of research in the scientific communication literature.

Lievrouw (1990) provides examples of research carried out within these four research programs and some of them will be mentioned in the following. Examples of artefact studies are the studies of the exponential growth of science by de Solla Price (1963) and the mappings of scholarly communication by Garvey and Griffith (she refers to Garvey, 1979). Examples of user studies provided by Lievrouw are studies of information use such as the studies conducted by Garvey and Griffith leading to the mappings and the user studies. The beginning of a user-centred orientation in information research is often related to the Dervin and Nilan (1986) review, although Talja and Hartel (2007) provide an overview of studies conducted between the 1950s and 1970s. Network studies are exemplified by the studies of invisible colleges by Crane (e.g. 1972) and Mullins' study of molecular biology (Mullins, 1972). Finally, lab studies are exemplified by the ethnographic study of a neuroendocrinology research laboratory by Latour and Woolgar (1979) and the study of biochemistry and microbiology research laboratories by Knorr-Cetina (1981, 1982).

3.2.4. Schema for studying social and cognitive processes in science

Griffith (1990) distinguishes between process, structure and context. Note that Griffith uses a different understanding of process and structure than Lievrouw (1990). He argues that:

> The goal of studying science is to seek an understanding of it as human behavior, not as technical content. For most investigators, the goal does not lie in using science to understand behaviour as behaviour or communication as communication. Modern science is too laden with complexity to serve well as a laboratory; the goal, instead, is to seek understanding of social and cognitive processes general to science (Griffith, 1990: 31)

Using the following schema Griffith (1990: 31) states that:

> [W]e are trying to understand the bottom six boxes through studying the top three. Communication is the only general scientific behaviour; other behaviours are mostly specific and technical. Information and its representations are its principal and general articfacts. Scholarly syntheses

are its final integrations and mechanisms for conservation of scientific achievement. These three are the principal data for understanding general social and cognitive processes in science.

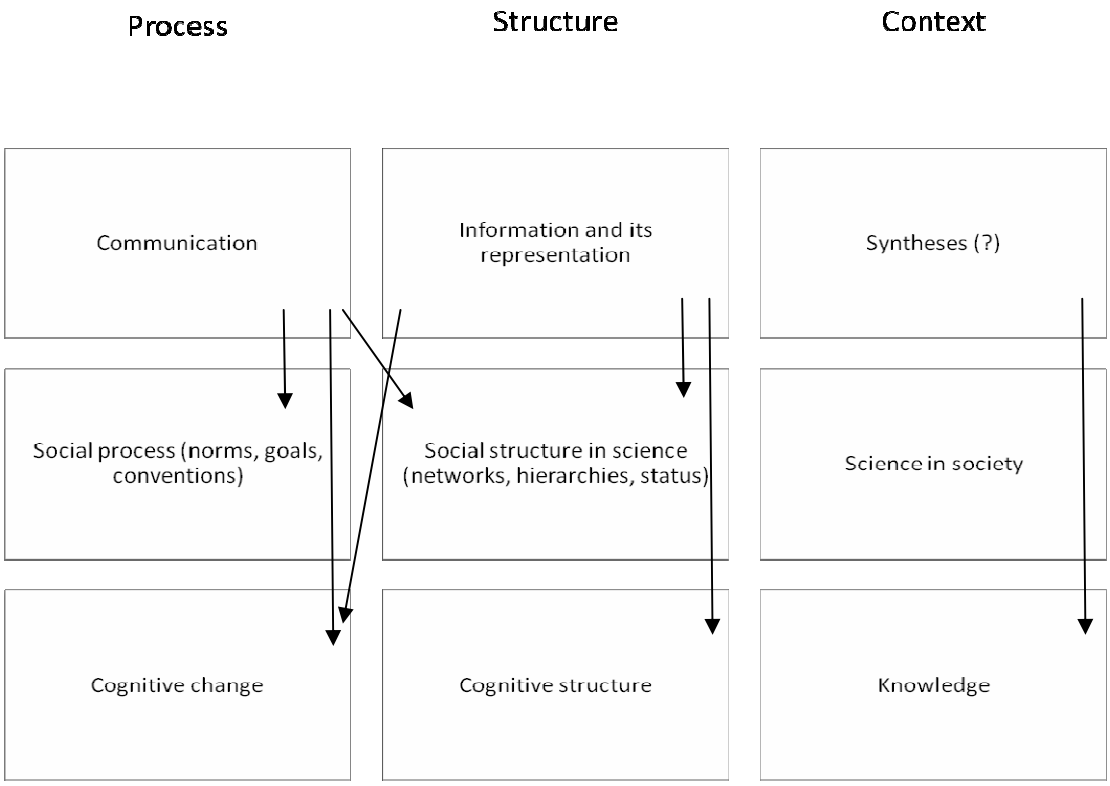


Figure 2. A schema for studying social and cognitive processes in science.

Note: The top three boxes indicate behaviours and articfacts used to analyze social and cognitive events in the bottom six boxes. From Griffith (1990: 32).

Griffith (1990) provides examples of how the model can be used to characterise research on scholarly communication. Merton (1942, 1973) is an example of research in social process and structure. Kuhn (1962) is an example of research in cognitive change and structure. The empirical studies by Price as well as Garvey and Griffith are mentioned by Griffith; however, not placed in the model. The contributions situated in

the model "were amplified by a number of [….] empirical findings and models for information". Consequently, the model by Lievrouw (1988) primarily characterises empirical work whereas the model by Griffith is developed primarily to characterise theoretical work.

3.2.5. Theoretical framework for the dissertation

Drawing on the models by Lievrouw (1988) and Griffith (1990) the theoretical and empirical work of particular interest is the work characterised by Lievrouw as artifact and user studies as well as the work characterised by Griffith as being related to process and structure.

More specifically, the framework for the present dissertation consists of theories of citing and mappings of scholarly, communicative processes. The latter is included as means to analyse any possible changes in the structure of scholarly communication. The former is included for the purpose of being able to analyse the far-reaching consequences of a potential open access citation advantage.

In relation to the citation theories a comment needs to be made as it is worth noting the difference between citing behaviour and information seeking behaviour. Availability as an attribute of the citing / cited pair is obvious to an extent that Borgman and Furner (2002) do not provide references to support the claim. However, it is worth noting that although availability in principle is required in order to cite a given document it may not necessarily be the case in practice. Authors may cite a publication that is not available to them or that is available but they have not read as suggested by Eichorn and Yankauer (1987), Evans et al. (1990), and Wright and Armstrong (2008). An argument also put forward by Latour (1987). Although read, the reference can be found in databases or in secondary sources (see e.g. Navarro (1999)). Authors do not necessarily read a paper before citing it but the authors must as a minimum locate a reference to the work (unless the reference is completely fraudulent). Consequently, the information seeking process is of great importance to the citing process.

Convenient access and ease of use have been shown to influence the choice of information channels (Hardy, 1986; Leckie, Pettigrew & Sylvain, 1996; Liu, 2006). A

recent study by Bronstein & Barachson-Arbib (2008) finds that four obstacles are encountered by scholars when searching for information and among these is distance (defined as when the item can only be found at a distant location). However, they also find a negative correlation between the difficulties in using an information channel and the reported frequency of use of that channel. Bronstein and Baruchson-Arbib (2008: 141) conclude that:

> Although the accessibility and ease of use of the information channel are still important parameters in the selection process […] scholars still choose the information channel based on the quality of the information provided by that channel.

Scholars may have difficulties using an information channel and their access is thus hindered but their choice and use of a specific channel is not influenced by these obstacles. However, whether or not they choose to cite the items found in their search for information is another matter. Readership does increase the chance of citing the publication as found in studies by Kurtz et al. (2005), Perneger (2004), Moed (2005) and Brody, Harnad and Carr (2006). According to Klamer and van Dalen (2002) the inflation of the number of publications produced by each scholar implies that the number of articles seeking attention is exploding. Attention is thus a scarce resource, and the inflation probably means that the competition for the limited attention becomes even harder. Readership is thus an increasing constraint on the chance of a given publication being cited.

Open access can be related to both information seeking processes and citing behaviour, and probably the latter follows the former process. However, it is of utmost importance that the two processes are not dealt with collectively as the information seeking process cannot be regarded to determine the citing process. The information seeking processes of scholars are not the main focus and consequently, theories and studies focusing on information seeking processes are not included. As a result of that discrimination, cost-benefit theories such as the optimal foraging theory and least effort theories are not included in the following despite being mentioned by Nicolaisen (2007) in a review of citation theories.

3.3. Citation theories

A possible increase in citations to open access publications due to the increased access is of great importance as it is deeply related to the understanding of what a citation represents. As Nicolaisen (2007: 633) states:

> [I]f we are to understand the nature of the citation, we need to understand the nature of the reference. And if we are to understand the nature of the reference, we need a theory of citing that explains why authors cite the way they do.

Consequently, should scholars cite publications primarily due to an availability criterion this will have bearing on every study relying on the use of references and / or citations for analytical purposes.

A number of different citation theories coexist and at times compete. Bornmann and Daniel (2008) as well as Davis (2009) reduce the citation theory debate to a question of two so-called camps. Davis (2009) refers to the two camps as being "reward" versus "persuasion". Bornmann and Daniel (2008) use a different terminology namely: "the normative theory" and the "social constructivist view" of citing behaviour which is also the terminology used by e.g. Nicolaisen (2007). However, Nicolaisen (2007) provides a much more elaborated overview of theories and studies of citing behavior.

The following is a brief overview of citation theories and their relation to the hypothesis of the open access advantage. The first theory analysed in relation with the open access citation advantage is the normative theory of citing, followed by the handicap principle as citing theory and finally, the social constructivist theory of citing is analysed. The starting point in the overview of each theory is the extreme case of a substantial open access citation advantage as the case in the study by Lawrence (2001) showing an advantage of about 150 percent. However, in the cases of the theories offering no room for a substantial open access citation advantage the case of a minor open access citation advantage is considered.

3.3.1. The normative theory of citing

The normative theory considers science a normative institution administered by internal rewards and sanctions. This theory was laid down by Merton who argued that science is bound by norms and values respected by the scientific community voluntarily. The norms are articulated in the form of prescriptions, proscriptions, preferences and permissions and legitimised in terms of institutional values. Merton (1942, 1973) identified four sets of institutional imperatives to be universalism, communism, disinterestedness and organised scepticism.

It is the interplay between the stimulus to conformity and assents against nonconformity that warrants science as a part of society via the institutional demands to publish. According to the normative theory the structure in science encloses obligations to cite the knowledge the work is build upon. Citations are objective as they are awarded according to the scientific method and without influence from personal factors.

> [T]here is the [….] obligation, within the institutional structure of science, for the user of that freely published knowledge to make open reference to the sources to which he is indebted […] (Merton, 1977: 49).

Consequently, in general citing easily accessible literature only for the reason of the easy access would only rarely allow the author to cite the sources to which he is indebted without requiring publications that are difficult to get access to. Such reference lists would tend to consist of a sizeable amount of sources referring to other sources to which the present author is also indebted to (but has left out because they where hard to get access to), secondary sources, un-refereed or even un-scientific sources. Obviously, such reference lists would not occur in peer reviewed publications because the reviewers would not allow it. The obligation to cite the sources to which the author is indebted thereby give rise to the opportunity to regard citations as appropriate for assessment of scholarly impact or excellence.

> [I]f one's work is not being noticed and used by others in the system of science, doubts of its value are apt to arise (Merton, 1977: 50).

The normative theory offers no room for increased availability as reason for citing. Obviously, the scholar would have to have access to a publication to be able to read, use and cite it. But easy or increased availability is no legitimate reason for citing a publication.

Following the normative theory a minor open access citation advantage can be ascribed to the differences between the norms and the actual behaviour of scholars. Merton argues (1942, 1973: 276):

> There is competition in the realm of science, competition that is intensified by the emphasis on priority as a criterion of achievement, and under competitive conditions there may well be generated incentives for eclipsing rivals by illicit means.

Consequently, incentives do exist that may lead an author to deviate behavoir. Zuckerman (1988: 520) has commented on the paradox that "while prescribing the terms of conformity, the normative structure can also generate deviance." Merton (1957, 1973) overviews a number of different forms of fraud: the most deviant behaviour in science. Merton (1942, 1973: 276) does, however, stress that:

> Cultism, informal cliques, prolific but trivial publications – these and other techniques may be used for self-aggrandizement [with reference to Wilson, 1942]. But in general, spurious claims appear to be negligible and ineffective. The translation of the norm of disinterestedness into practice is effectively supported by the ultimate accountability of scientists to their compeers.

Summing up, the cases of deviate behaviour are negligible, however, they do occur. Consequently, the theory does have room for a small open access citation advantage; although it can never be substantial.

3.3.2. The handicap principle as theory of citing

The handicap theory or theory of costly signalling was developed over a number of years. The work leading to the handicap principle started as an attempt to explain the evolution of the peacock's tail and its broad applications - including human behaviour - were developed later on (Zahavi & Zahavi, 1997: XVI). The handicap theory is based on a simple idea (Zahavi & Zahavi, 1997: 229):

> [W]aste can make sense, because by wasting one proves conclusively that one has enough assets to waste and more. The investment – the waste itself – is just what makes the advertisement reliable.

The peacock's tail is the classical example of sexual selection producing some trait despite its being disadvantageous to biological fitness. In this case of sexual selection only the quality signallers can afford to produce extravagant waste and thus send a handicapped signal. As Zahavi (1975: 207) expresses it:

> An individual with a well developed sexually selected character is an individual which has survived a test. A female which could discriminate between a male possessing a sexually selected character, from one without it, can discriminate between a male which has passed a test and one which has not been tested. Females which selected males with the most developed characters can be sure that they have selected from among the best genotypes of the male population.

According to the theory, costly behaviour is the key to reliable signals since the cost of a signal guarantees its honesty. Zahavi (2003: 4) stresses that a handicap does not evolve to decrease fitness. He argues that:

> The selective process by which individuals develop their handicap increases their fitness, rather than decrease it. If "cost" is measured by a loss in fitness, then handicaps do not have a cost for honest signalers, since honest signalers increase their fitness by signalling.

Due to criticism of the model of the handicap Zahavi (1977b: 603) stressed that the relationship between quality and the handicap may be more sophisticated [than a simple additive value to the handicap]. The investment in a handicapped signal is not necessarily very high as the investment is proportional to the potential gain. In 1997 he describes his response to his critics:

> An honest threat communicates reliably one's ability and willingness to fight. A reliable threat leaves the signaller open to attack. This increased risk is acceptable to the honest signaller – the one who thinks the objective is worth a fight, and that it can win against a particular opponent. Such a threatener has already decided to fight if its opponent does not retreat, and the increased likelihood that it will actually have to fight does not deter it. A bluffer – one who tries to gain by threats alone but is not really willing to back up its threats by fighting – would find the increased likelihood of being attacked too risky (Zahavi & Zahavi, 1997: 16).

The theory has been applied broadly in a number of disciplines (the reader is referred to Nicolaisen (2007) for a short overview), however little in information science. An analogous theory to the handicap principle not stemming from evolutionary biology exists. This theory is more likely to have been applied in information science as the theory was introduced to the field in 1974.

In economics signalling indicates that one party (termed the agent) conveys some meaningful information about itself to another party (termed the principal). Probably preceding the work by Zahavi, Spence (1973) introduces indices and signals in economics using a terminology suggested by Robert Jervis:

> Indices are attributes over which one has no control, like gender, race, etc. Think of them as unalterable attributes of something, not necessarily a person. Signals are things one does that are visible and that are in part designed to communicate. In a sense they are alterable attributes (Spence, 2002: 434).

Spence (1973) illustrates the definitions and properties of signalling equilibria using the case of education in the job market context. Spence (2002: 436) describes the model:

> The idea behind the job market signaling model is that is there are attributes of potential employees that the employer cannot observe and that affect the individual's subsequent productivity and hence value to the employer on the job. [….] Now suppose that there is something called education, which we will denote by E, that can be acquired or invested in. It is assumed to be visible, and its acquisition costs differ for the two types.

There are two components in model that needs to be consistent. The first is that individuals make rational choices regarding their investment in education and secondly, employers have beliefs about the relation between the signal and the individual's underlying productivity. This implies that the beliefs of the employers determine the wage they offer individuals with varying levels of education.

> These wage offers in turn determine the returns to individuals from investments in education, and finally, those returns determine the investment decisions that individuals make with respect to education, and hence the actual relationship between productivity and education that is observed by employers in the marketplace. This is a complete circle. Therefore it is probably more accurate to a say that in equilibrium, the employers' beliefs are self-confirming (Spence, 2002: 437).

The example of education as a signal in the job market implies that from the perspective of the handicap theory education becomes a costly signal only affordable to the strong candidate whereas education at the same level is too costly for the weaker candidate considering the costs. Consequently, the theory is analogous to the handicap theory. Spence concludes in 1974:

> Effective signals tend to be those that are somehow self-validating. Many effective signals are costly to the sender, but more costly if the message is false. Thus, guarantees can signal product durability because, if the product is not durable, it is very costly for the manufacturers who issue the same guarantee (Spence, 1974: 74).

Spence (1974) introduces signalling models to the field of information science. He presents the models as attempts to model the information transfer process in markets (Spence: 1974: 62). A search for citations to Spence within the field of information science reveals that he is typically being cited in information science for the economist's view of information and in studies of commerce. An example is the work by Afzal, Roland and Al-Squri (2009) that cites Spence (1973) to describe the behaviour of consumers and firms in a given market. Spence is not being cited for providing an opportunity to explain the behaviours of scholars.

Within information science Nicolaisen (2004, 2007) has discussed the potentials of the handicap theory, and Nicolaisen and Frandsen (2007) exemplify the potentials of the handicap principle on the practice of academic book reviewing. Nicolaisen (2004) likens references to threat signals and argues that the references are a sign of confidence.

> A stack of references is a handicap that only the honest author can afford. […] Modalized references expose themselves like the vocalization of a bluffing threatener. A skilled rival will detect the false sound right away and then know where to attack. The potential cost of making such a sound will often make the author reconsider his deceitful behavior Nicolaisen (2004: 85).

Consequently, a reference list characterised by a great number of open access publications that are primarily cited due to the easy and convenient access it would be a handicapped signal only few authors can afford to send. As Nicolaisen (2004: 85) puts is: "Unconfident authors would usually not dare to risk the potential loss of reputation." This implies that the handicap principle as citing theory does not allow a large open access citation advantage. The handicap principle can, however, accept a small open access citation advantage. Nicolaisen (2004) uses the handicap principle to explain the standard account in citation theories and any open access advantage would be due to rarely occurring, dishonest signals. Nicolaisen (2004: 85) stresses that not all references need be honest signals and he proposes that:

> [T]he handicap principle secures that citing authors credit their inspiration and sources in an honest way, to a tolerable degree – enough to save the scientific communication system from collapsing.

Consequently, too much "cheating" would disrupt the correlation at the foundation of the system, causing it to collapse. Zahavi's handicap principle provides a possible mechanism by which a signalling system could be kept honest. If growing a long tail was not a prohibitively costly signal for the weak peacock, all peacocks would grow long tails and tail length would no longer be an honest signal. Only communication systems in which the signal is costly can escape collapse. Consequently, the stability of a communication systems depends on the how costly the signals are.

This may be a potential weak point in the use of handicap principle to explain the communication behaviour of scholars. We do not know the exact cost of a signal nor the exact potential gain. Consequently, we cannot determine if the system as we know it is about to collapse caused by "deceiving" behaviour such as a substantial open access advantage would imply. This brings us in an unfavourable situation where the theory both can and cannot accept an open access citation advantage.

3.3.3. Social constructivist theory of citing

There are marked differences in the conceptions of the normative theory and the social constructivist viewpoint. Knorr-Cetina (1995: 141) describes the emergence of the social constructivist viewpoint as "a thoroughgoing sociological contextualization of science" (Knorr-Cetina (1995: 141).

> Constructionism holds reality not to be given but constructed: It sees the whole as assembled, the uniform as heterogeneous, the smooth and even surface as covering internal structure. There are, for constructionism, no initial, undissimulatable "facts": neither the domination of workers by capitalists, nor scientific objectivity, nor reality itself (Knorr-Cetina (1995: 147).

As we can see Knorr-Cetina uses the term constructionism whereas constructivism is used in other writings (e.g. Knorr-Cetina, 1981). Constructivism has been used in various contexts and with different meanings in the literature but Sismondo (1993) provides a detailed account of the differences. Sismondo (1993) identifies four prominent interpretations of "social constructions". He argues that they are typically merged in science studies although they are easy separable and should be kept separated due to variance in tenability. The four interpretations are (Sismondo, 1993: 547):

1. An interpretation in terms of large social projects, whereby such things as cities, economies, legislation and knowledge are constructed by many people interacting, possibly with differing or conflicting goals.
2. A geometrical interpretation, whereby conceptual entities are constructed given some fixed points (data, tools, resources and the like).
3. A more physical interpretation, whereby new material objects are made form other ones.
4. An interpretation in which objects are constructed out of world views.

The fourth interpretation Sismondo finds less plausible due to weak arguments supporting it. Although Sismondo argues for keeping the interpretations separate, he nonetheless concludes that:

> Construction metaphors in science studies have provided valuable insights, many of those consequences have yet to be explored. For example, the laboratory in all its artificiality, the third type of constructed reality, is something at which more people have been looking in recent years. The insight that it *is* constructed remains to be systematically thought through and incorporated into our picture of scientific activity (Sismondo, 1993: 547).

He sums up by arguing that a simple recommendation can be made on the basis of the number of valuable insights provided by social constructivism: "we should more often recognize the existence of social objects in science" (Sismondo, 1993: 548). Consequently, he treats them as a whole.

Following Collin (1997), Nicolaisen (2004, 2007) makes a separation between two distinct groups of scholars labelled social constructivists and disintegrates the social constructivists (e.g. Emile Durkheim, Peter Berger, Thomas Luckmann, Don Zimmerman, Melvin Pollner, Alisdair MacIntyre, and Peter Winch) and the science constructivists (e.g. Barry Barnes, David Bloor, Michel Callon, Harry Collins, Karin Knorr-Cetina, Bruno Latour, and Steve Woolgar). Nonetheless, he uses the term social constructivists to label the latter group. Zuckerman (1988) uses the term constructionists to label the same group of scholars as Nicolaisen. She identifies four variants of constructionist studies of scientific knowledge and characterises them by the names used by the analysts themselves:

1. Relativist studies (e.g. Bloor, 1976; Collins, 1981)
2. The interests model (e.g. Restivo & Lauglin, 1987)
3. Constructivist studies (e.g. Latour & Woolgar, 1979; Knorr-Cetina, 1981)
4. Discourse and text analysis (e.g. Mulkay & Gilbert, 1982; Gilbert & Mulkay, 1984; Mulkay 1985; Bazerman, 1984, 1985)

Zuckerman stresses that the four variants of studies under the constructionist heading draw on different intellectual antecedents, however they share two philosophical precepts.

> First is the idea that facts are theory laden, that what we take to be evidence is shaped by our theories and their constituent concepts, other related conceptual schemes, and by our ideas about observation and measurement—a view often attributed to Kuhn and Feyerabend but of course older and not necessarily associated with a strictly relativist perspective. The second is that scientific theories are undetermined by evidence (generally called the Duhem-Quine thesis, but sometimes the Duhem-Quine-Hesse thesis). Many theories, it is said, can fit the same facts or, put another way, empirical evidence does not provide firm grounds for accepting or rejecting theories. If theories cannot be rejected or accepted on the grounds of the evidence brought to bear on them, then scientists have a good deal of leeway in theory choice. Zuckerman (1988: 547).

Consequently, the four variants of studies identified by Zuckerman can be treated as a whole. The terminology herein will follow that of Nicolaisen (2004, 2007) using the term "social constructivist theory".

Social constructivist theory of citing rejects the assumption that citations represent what the present work is influenced by. On the contrary, scholarly publications can be considered fiction, perversion or storytelling and an element in the persuasion of readers is the citations. Citations are one rhetorical device that scientists use in order to provide support for their claims (MacRoberts & MacRoberts, 1996).

According to both Borgman and Furner (2002) as well as Bornman and Daniel (2008) availability of publications is one of the many factors the probability of being cited is depending on. Availability forms an argument for the social constructivists in their critique of the normative theory (Bornmann & Daniel, 2008: 46).

Writing a paper is detached from the actual research which implies that an author subsequent of doing the research look for publications that fit the author's purpose. The author place citations in their publications that can persuade the audience and should the author wish to persuade the audience that he or she is a confident user of the internet as source for scientific information then exaggerated citations to open access publications can be the means. Consequently, the social constructivist theory of citing allows for both a minor and a substantial open access citation advantage. The social constructivist theory is thus the only citation theory offering a framework allowing a substantial open access citation advantage.

3.3.4. Summing up

The debate on open access citation advantage is primarily taking place on the basis of empirical data and personal opinions. However, it should be clear from the overview of citation theories that an open access citation advantage has bearing on our understanding of citations. All citation theories can allow for a minor citation increase to OA publications. However, only the social constructivist theory of citing can allow for a major citation increase. We should bear in mind that an information seeking process often precedes a citing process and consequently, it would be difficult not to

expect some consequences of the former for the latter. Accepting a major open access advantage implies accepting reasons for citing not allowed by the normative theory of citing as well as the handicap principle as theory of citing.

Concluding on the studies of the thesis regarding citing behaviour will include the implications of the results for citation theory.

3.4. Mappings of communicative processes

A number of different models exist that conceptualise scholarly communication. They are characterized by varying objectives and foci.

3.4.1. The Information chain

The information chain is also known as information flow, information transfer and can be defined as "the institutional and documentational structure of human communication" (Duff, 1997: 179). The model emphasises publications and related documents (see e.g. MacKenzie Own & Halm, 1989; Meadows, 1991; and Duff, 1997). Figure 3 is an illustration of the simplest information chain model as argued by MacKenzie Owen (2002) who explains that the information chain model identifies:

> "[S]pecific roles (such as knowledge creation, publishing, archiving, intermediation and use), and actors that perform these roles (such as researchers or research institutes, publishers, libraries and users)."

A different terminology, but essentially the same structure, is suggested by Garvey and Griffith who identify four general processes of scientific information exchange: "the origination of information, its transmission, its storage, and its use" (Garvey & Griffith, 1963: 1). They relate three roles to the aforementioned four processes: "the source, the user, and information conveyance devices that have both transmission and storage functions" (Garvey & Griffith, 1963: 1).

Figure 3. The traditional information chain model (MacKenzie Owen, 2002).

The model focuses on a chain of information functions through which knowledge is transmitted from the originator to the user or from producer to consumer. The user or consumer would primarily be peers which is also stressed by Garvey, Tomita and Woolf (1974: 115) who argue that the largest group of users of scientific information is the also the producers of that knowledge. However, the model can be applied widely, including the dissemination of non-scholarly publications (Duff, 1997). The actors fill specific functions within the context of their respective roles and define the form and substance of their responsibilities. The functions of publishers can be defined in accordance with the task performed as e.g. editing, printing, marketing, distribution or the functions can be defined more conceptually as e.g. dissemination, quality control (MacKenzie Owen, 2002).

Several models based on the information chain exist (see Duff, 1997). The models are represented in terms of flow of information or cycles. An example of an information chain model is the UNISIST model of scientific and technical communication (UNISIST, 1971) whose strength is that the model enables a distinction between primary, secondary and tertiary sources and services. The model, however, focuses less

on the scholarly communication taken place via the various sources which is the case with most of the existing models stemming from the information chain. Regazzi and Caliguiri (2006: 183) even refer to the information chain model as the "publishing supply chain". Consequently, the majority of the models focuses on the artefacts themselves and less on the scholarly communication functions they serve. An exception and probably the most well known model of scientific communication stemming from the information chain is the Garvey and Griffith model of scholarly communication that was presented more than forty five years ago (Garvey & Griffith, 1963).

3.4.2. The Garvey and Griffith model of scholarly communication

Their objective was, as already mentioned, an effort to describe the scholarly communication in psychology by identifying people and institutions associated with the roles in the information chain model (Garvey & Griffith, 1963: 1). However, their second objective led to the famous Garvey and Griffith model of scholarly communication. They intended to develop:

> "[A] description of the time characteristics of each process and of the "filtering" that determines the amount and type of information which will be available to the user (Garvey & Griffith, 1963: 1).

They set out to map the communication processes of researchers and the project was not intended to focus on a limited set of media or functions (Garvey & Compton, 1967). The project on scientific information exchange in psychology was established in 1961 by the American Psychological Association, although aided and assisted by the National Science Foundation (Garvey & Compton, 1967). Initially, the model was based on their observations within the field of psychology but was later found to be useful as a general model of scholarly communication (e.g. Hurd, 1996). From their empirical work emerged, using the words of Belver Griffith, "the natural history of the production, dissemination and use of information" (Small, 1997). The model emphasises the timeline for scholarly communication. They outline the stages and provide details of the stages in psychology from initiation of the research through publishing a journal article to having the work integrated within the context of other works in a review (Garvey &

Griffith, 1964). According to Griffith (1990: 40) their work made three general contributions:

> (a) the development of a model for disciplinary processing of information, (b) the important distinctions in function and structure between the formal and informal realms of information dissemination, and (c) the identification of the role of the productive scientist as driving and controlling scientific exchange.

The work by Garvey and Griffith is based on a number of studies and consequently a number of different data sets. The American Psychological Association's Project on Scientific Information Exchange in Psychology reports are the joint products of the staff of that project, directed by Drs. Garvey and Griffith from 1961 to 1966, and then by Dr. Griffith until 1969. An overview of the data forming the basis of the work by Garvey and Griffith is organised according to the report numbers in appendix 1. However, it is worth noting that their famous model is published as early as August 1963 at which point in time only 7 of the 21 reports were available (Garvey & Griffith, 1963: 11). Consequently, it is not unlikely that the last of the reports could be described as elaborating on some of the elements of the model more than contributing to the development of the model. Furthermore, it is also worth mentioning that particularly the most recent reports are based on larger samples than the early ones.

Garvey and Griffith state that the model is based on empirical data, although it is difficult to determine how and what data has been used. They state that for reasons of convenience:

> [T]he process is described in terms of the average producer of a research report in a 'core' psychological journal; the time intervals are actually medians of the collected data in one or more studies (Garvey & Griffith, 1963: 15).

These limitations of their work should be kept in mind; nonetheless the model provides an excellent and straightforward overview of scholarly communication in psychology. The four ovals in figure 4 and 5 represent a significant phase in the work of the researcher. The rectangles to the right are the possible forms of oral reports. To the

immediate left we find the possible forms of written reports and to the far left we find secondary publications. The figure is shaded in selected areas to indicate the information generally available to the scientific public and the information only available for a restricted audience (Garvey & Griffith, 1963).

The average time period from initiation to journal publication is about 3 years and the time period from journal publication to review is about 2 years. Incorporation in specialized texts and treatises is described to be 13 years after the work is initiated (Garvey & Griffith, 1972). The following figure is a diagram depicting the process from initiation of the work to the integration of the findings into the fund of scientific knowledge. Following the original figure from 1963 is a version of the model from Borgman (2007) that is reprinted much clearer.

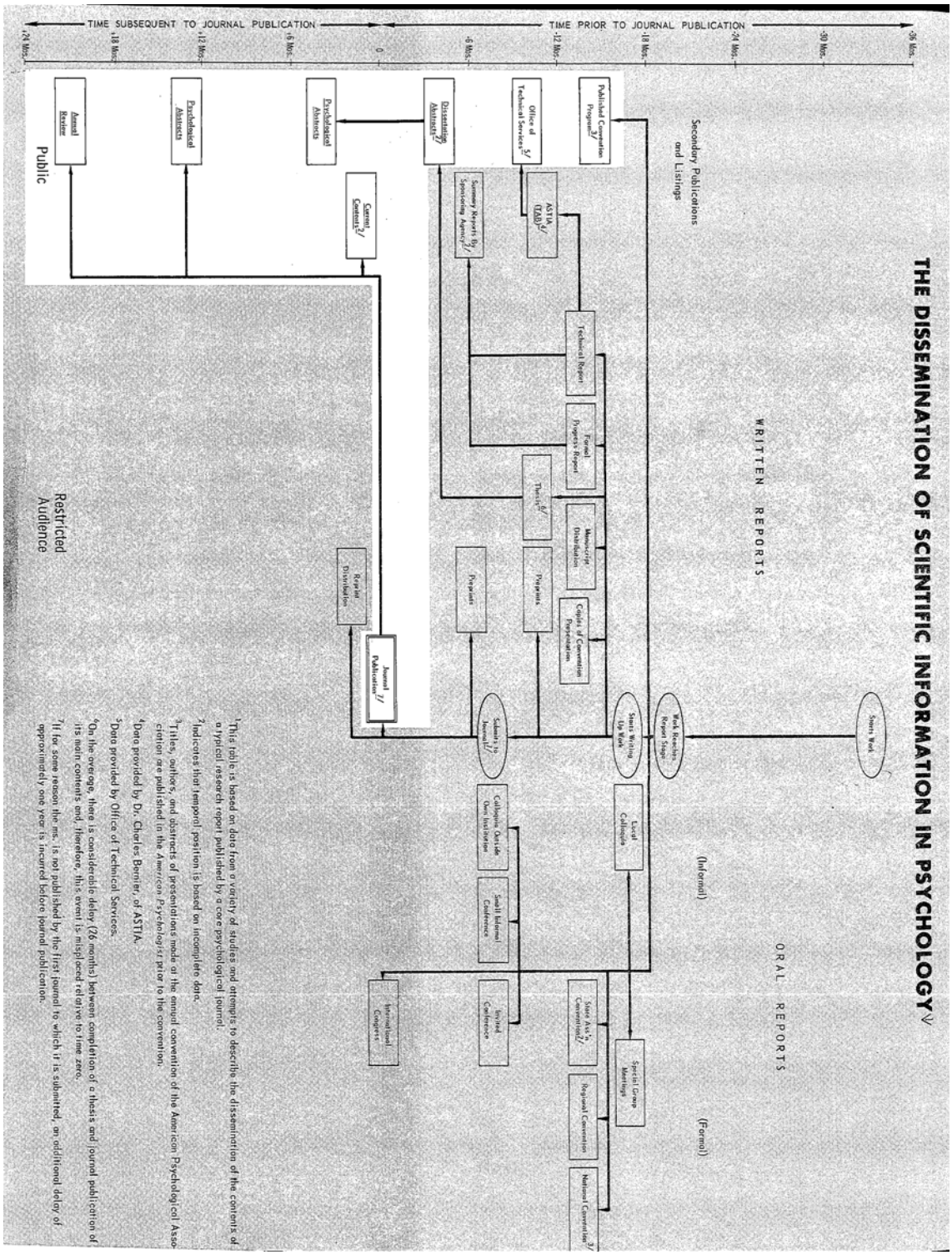
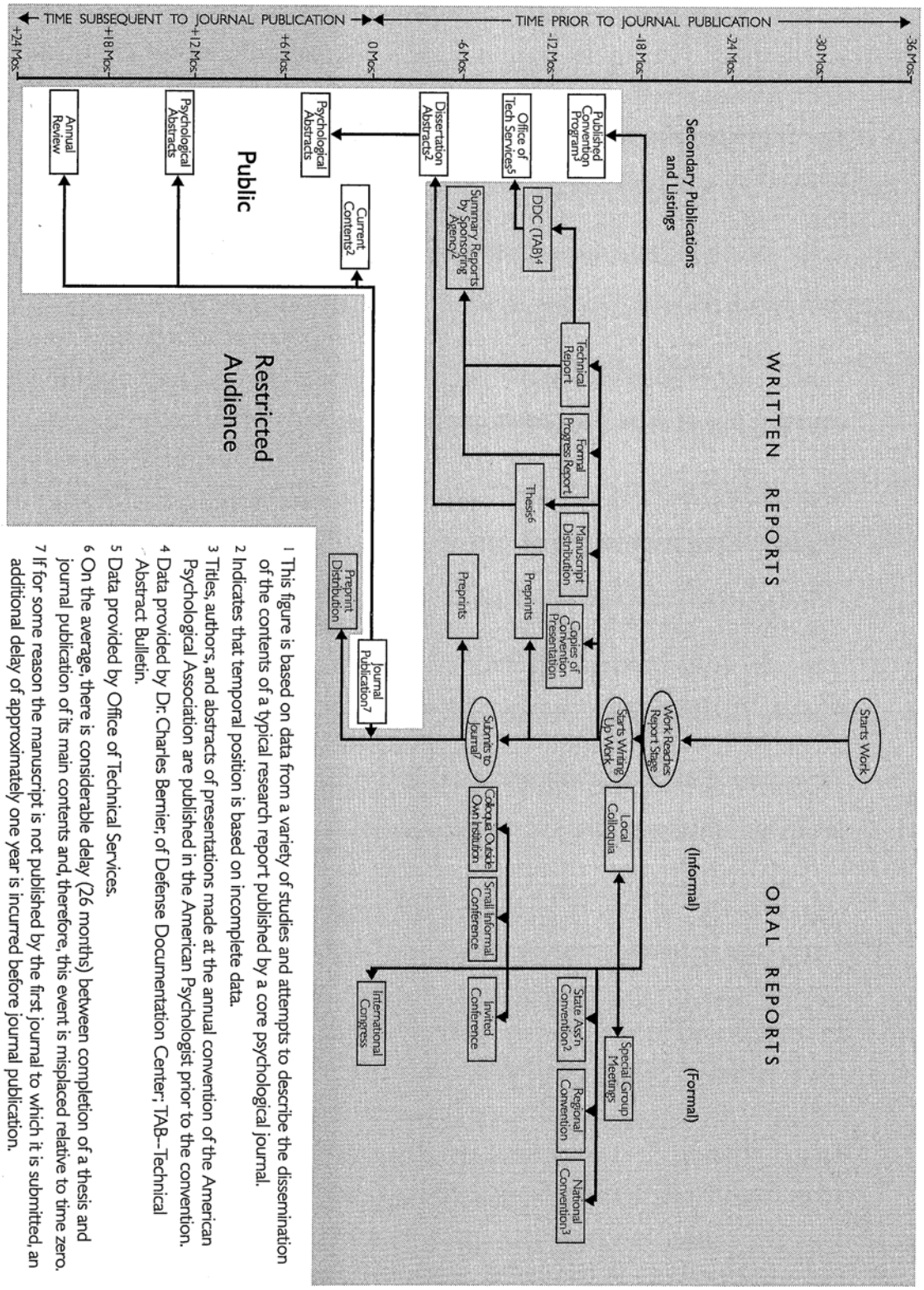


Figure 4. The dissemination of scientific information in psychology (Garvey & Griffith, 1963).



1 This figure is based on data from a variety of studies and attempts to describe the dissemination of the contents of a typical research report published by a core psychological journal.

2 Indicates that temporal position is based on incomplete data.

3 Titles, authors, and abstracts of presentations made at the annual convention of the American Psychological Association are published in the American Psychologist prior to the convention.

4 Data provided by Dr. Charles Bernier, of Defense Documentation Center; TAB—Technical Abstract Bulletin.

5 Data provided by Office of Technical Services.

6 On the average, there is considerable delay (26 months) between completion of a thesis and journal publication of its main contents and, therefore, this event is misplaced relative to time zero.

7 If for some reason the manuscript is not published by the first journal to which it is submitted, an additional delay of approximately one year is incurred before journal publication.

Figure 5. The dissemination of scientific information in psychology (Borgman 2007: 50).

As seen in the figure Garvey and Griffith place significant value in the informal communication during the entire period from initiation of the work to journal publication. As stated by Garvey and Griffith (1964: 1655):

> [T]he literature is only a portion of a system that encompasses many forms of scientific exchange; and without denying the great importance of the archival journals, it may be said that they have received a disproportionate share of the attention being given to the mechanisms by which scientific information is disseminated.

Garvey and Griffith (1964: 1655) describe the non-journal communication as:

> [T]he exchange of new scientific information between its principal producers and consumers does not wait upon journals. The active scientist makes use of a whole network of means of communication, many of them informal or of a small range, and yet apparently highly efficient. Their efficiency lies not only in their expeditiousness but also in their selectivity, for the group that is actively interested in a particular set of findings is often quite small.

Informal communication takes the form of reports to small audiences, reports to sizeable restricted audiences, preprints of technical reports, preprints of manuscript, seminars, meetings and conferences (Garvey & Griffith, 1972). Garvey (1979: 23-25) describes the informal domains of communication structures as ephemeral, lacking stability, genuinely interactive and redundant. There is considerable redundancy in the overall system of scientific communication. Garvey and Griffith (1967) state that the formal channels contribute less to the redundancy because of the norm of not publishing the same research in different outlets but "it is not uncommon to find the same material repeatedly reshaped in various informal media, to fit the characteristics of the channel and the needs of the audience" (Garvey & Griffith, 1967, p.1013).

Furthermore, Garvey and Griffith state that the scientist "interacts with his immediate colleagues, usually to formulate his research problem more precisely and obtain specific information relative to methodology, apparatus, anticipated data analysis etc." (Garvey & Griffith, 1972: 128). The three major ways for scientists working within the same subject matter area of receiving useful information from pre-publication reports are by: the scientist had some pre-publication acquaintance with the specific research in the article, the scientist acquired useful information from a prepublication source (not surprisingly, a high share receives it from a pre-publication source) and face-to-face reports (Garvey & Gottfredson, 1977).

Garvey, Lin and Nelson (1970, 1971) examine the differences across disciplines. They conclude that:

> [T]here are striking similarities in the communication processes and practices of various disciplines, there are also some major differences among them which are important (Garvey, Lin & Nelson, 1971: 271).

Garvey, Lin and Nelson (1971) argue that although the communication systems in various disciplines have similar structures and the systems consist of similar elements, there are marked differences in the way the elements function, are being used and are interrelated. They also argue that lags and filtering differ in loci and extent across disciplines. This is illustrated by figure 6 that illustrates the dissemination schedule of information contained in journal articles within the social sciences and physical sciences (Garvey, Lin & Nelson, 1971: 260). The figure clarifies that there are different schedules in the social sciences than in the physical sciences (typically there are about 6 months more from initiation of the work to the journal publication in the social sciences than in the physical sciences if the manuscript is rejected the first time submitted). The timeframes may appear to be different now than thirty years ago. The diagram also points to where in the schedule the longer lags in the social sciences are to be found. Garvey (1979) argues that typically the lags are not due to the reviewing process but the fact that journals accept more manuscript every year than the editor can fit into the available pages in journal issues.

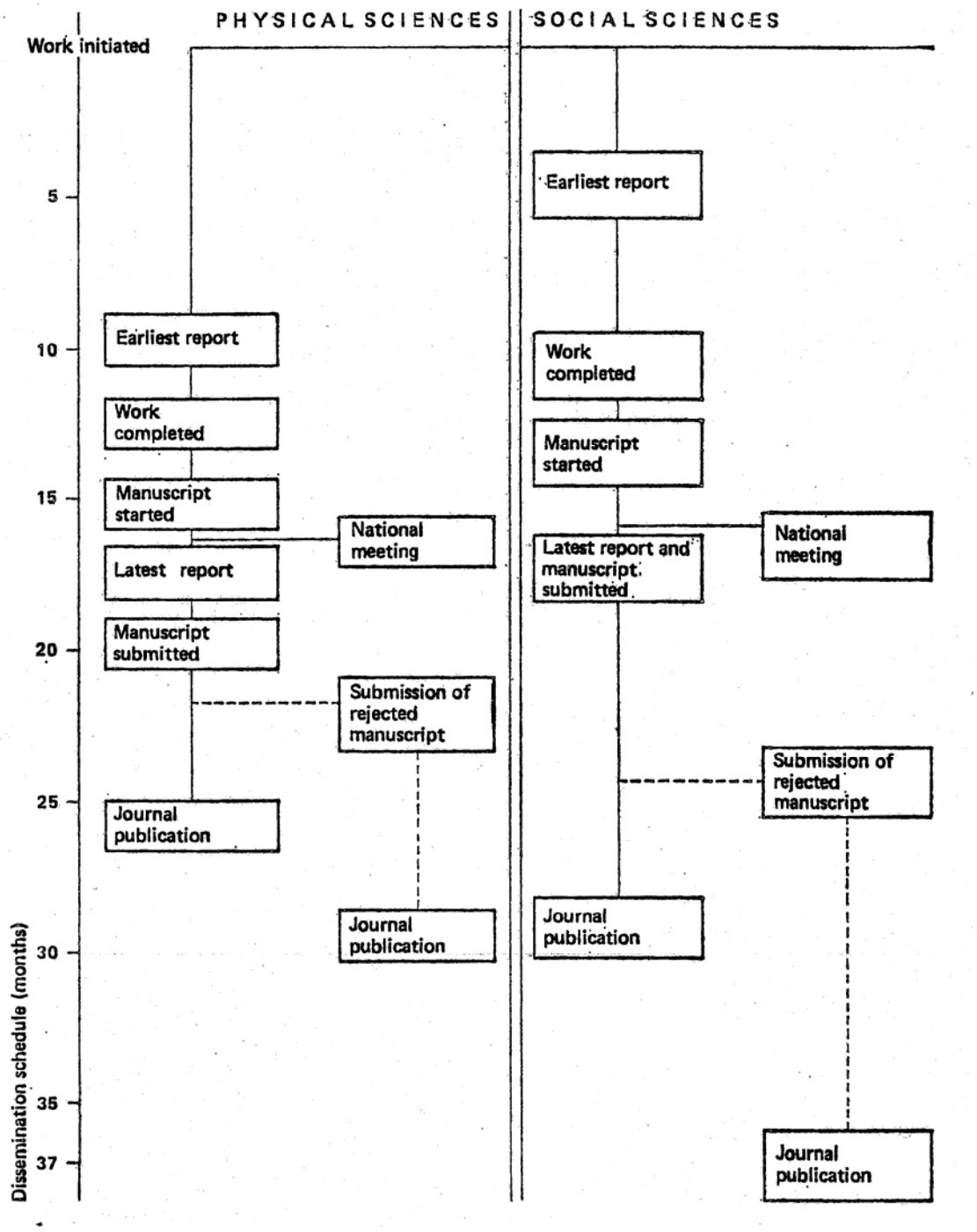


Figure 6. The dissemination schedule of information contained in journal articles (Garvey, Lin & Nelson, 1971: 260).

Garvey, Lin and Nelson (1971) find further differences across disciplines. Generally they find redundancy in each communication system (the same information being disseminated on a variety of occasions) however; they find that the extent and the purposes of the redundancy within a communication system vary. Furthermore, they find different behavior of scientists within different systems. They argue that it is due to "the differences in organization of the systems (and the different roles played in them by inherent elements" (Garvey, Lin & Nelson, 1971: 271). Finally, they find a highly interactive process within each communication system among informal and formal elements. The informal elements supplement the formal elements, and the efficiency of informal elements determines the degree and use of formal elements. Garvey, Lin and Nelson (1970, 1971) thus find that studying nine different disciplines reveals many differences in the communication process but they also find striking similarities.

A number of extensions and updates of the Garvey and Griffith model has been presented in the literature. Examples of extensions and updates are Hurd (1996, 2000, 2005) and Spink, Robins & Schamber (1998). Hurd (1996: 10) "examines the role of emerging information technologies and explores how these may catalyze changes in the communication system". She presents a series of models using the Garvey and Griffith model as starting point (a modernized, a no-journal, an unvetted, and a collaborator model). Spink, Robins & Schamber (1998) propose an extension to the model to include book reviews and feedback loops. Hurd (2000) describes a model for a future scientific communication which must be characterized as a contribution to the debate on current scholarly communication. Hurd (2005) presents an update of the model that incorporates developments in particularly document types and databases. All contributions to update or expand the model leave out one of the great strengths of the Garvey and Griffith model: the timeline based on empirical data. Furthermore, working with altering the original model tends to focus on artefacts and less on the communicative functions fulfilled by the artefacts. Consequently, the original model serves better as an analytical tool of the developments in scholarly communication.

3.4.3. Summing up

Numerous mappings of scholarly communication exist. They have different objectives and foci. The strength of the UNISIST model is that it enables identification of sources and the strength of the information chain model is that it enables identification of roles and functions of these roles. Probably, the most well-known model is the model of scholarly communication is the model by Garvey and Griffith. They modelled their mapping over empirical data. Although it can be difficult to determine how the data was used the strength of the model lies in the thorough work underlying the model and apparent simplicity of the model. Furthermore, their model enables identification of the communication process from an author perspective. The model is consequently the best applicable as analytical tool to investigate the implications of changes in scholarly communication.

3.5. Scholarly communication and open access

The present chapter aims at analysing the effects of open access on scholarly communication. For that purpose scholarly communication is defined and methods of analysing scholarly communication presented. The theoretical framework used to analyse changes in scholarly communication caused by an increase in open access publications is citation theories and mappings of scholarly communication processes. The former can cast light on changes in the citing behavior of scholars caused by open access, the so-called open access citation advantage. The latter can shed light on whether open access causes changes in the fundamental structures of communication identified more than forty five years ago by Garvey and Griffith. They actually recommended changing the scholarly communication to optimise the communication processes. Garvey and Gottfredson (1976) describe what they call a delayed-integrative journal system:

> These future journals would no longer accept piecemeal articles. Rather, they would require that such articles' publication be delayed until a coherent series of research works can be synthesized into a single major article.

> Journals would then serve mainly the functions of integrating and storing information in the archives of science (Garvey & Gottfredson, 1976: 170)

They suggest supplementing the delayed journal system with an element in the system that allows a "variety of forms of manuscripts, publish monthly abstracts of accepted material, and make the material available on an individual basis (Garvey & Gottfredson, 1976: 175). Consequently, their proposal is not that far from various open access initiatives. However, a crucial element in their proposal is that they remove some of the functions fulfilled from the journal and placed the functions elsewhere. It may seem that open access initiatives to some extent resemble their proposal without redistributing functions.

The next section provides summaries of the attached articles and following that are the analyses of the results of the articles seen in the light of the theoretical framework presented here.

4. Summary of the articles

The following includes summaries of the attached journal articles. They all contribute to answering the overall research question: What are the effects of open access on scholarly communication? The research question is being investigated on several levels and the dissertation consists of three levels of analyses. The first level consists of two studies of a general, more explorative character. The next level of analysis consists of two specific studies that both point back to the aforementioned tales of OA. The third level consists of a concluding, perspectival study. The summaries of the articles are presented in the order of the levels.

The first study deals with open access as an alternative database and to what extent open access based resources as a database reproduces the existing databases with an available and consistent indexing policy. The study treats the following question: How does OA affect the visibility of research? The second study focuses on the assumption that open access is a particular science phenomenon. The study deals with the following question: Is full OA coverage just a matter of time – how can the integration of open access in the existing publishing system be described in the sciences that are the leaders in establishing OA journals?

The third study looks into the assumption that OA is of greatest benefit to the developing countries. The main research question of the study is: What are the implications of OA for the developing countries in terms of using OA available research and publishing their own research? The fourth study deals with the most controversial OA tail: the open access citation advantage. The question is: What are the implications of the so-called open access citation advantage for publication types that have received little attention before the Internet?

Finally, the thesis includes a study of more perspectival character which tries to open up new perspectives within the field of OA and deals with the following question: What can open access provide of new possibilities in terms of research of scholarly communication?

The results of the articles are analysed in relation to the theoretical framework in section 5.

4.1. Intra-disciplinary differences in database coverage and the consequences for bibliometric research

Tools enabling citation analysis based on open access resources are considered an alternative to the existing citation databases (Noruzi, 2005; Bakkalbasi et al., 2006; Kousha & Thelwall, 2006; Neuhaus & Daniel, 2008). Examples of such tools are Google Scholar and Citeseer. Unlike the citation indexes, open access based resources do not allow systematic analyses of neither the indexing policy nor the consequences of it. The indexing policies of services based on open access resources are difficult to analyze. This is partly caused by a lack of available information of the indexing policy (e.g. Google Scholar) and partly because the service providing access to the data is not in control of the indexing policy (e.g. archives based on authors self-archiving their work). Consequently, the influence on bibliometric studies of the indexing policy is even more difficult to investigate as we are limited by the options available through the services and as Neuhaus and Daniel (2007) state:

> Google Scholar currently processes its sources in an unsystematic, unpredictable and fragmentary manner. For lack of adequate options for browsing, searching and saving results in structured output formats it is difficult to make even elementary bibliometric analyses efficiently.

However, the coverage is of great importance. The value of a specific database depends to a large extent on the coverage of the discipline(s) under study. According to Jacso (1997) the problem of coverage has both quantitative and qualitative aspects. The quantitative aspects concern among other things the size of the database(s), indexed document types, the number of English-language and foreign-language source documents, geographic coverage, and the time span and currency of the database(s). The qualitative aspects are partly about the inclusion of core journals and prestigious non-journal sources. A number of studies have determined the coverage of databases in

specific disciplines focusing on inter-disciplinary differences. However, little is known about the potential existence of intra-disciplinary differences in database coverage.

The article focuses on the coverage of bibliographic databases and the consequences for bibliometric research with a narrow focus on the disciplines of economics and psychology. Both are classified as well-covered in Moed's classification system of the ISI citation indexes coverage (Moed, 2005: 138), however, the article provides a more detailed analysis which is done by dividing the disciplines under study into research traditions and specialties.

Meadows (1998) characterises a specialty by the phenomenon or phenomena, which members of the specialty study. A research tradition is held together by common ontological assumptions about the nature of the world and methodological principles about how to revise theories and develop new ones (Laudan, 1977). Research traditions are consequently not the same as specialties.

The results document significant differences in the disciplines of economics and psychology, and revealed quite uneven coverage of economic specialties and psychological research traditions. The latter results are depicted in the following two figures. Table 1 shows the relative sizes of the four specialties vary considerably when using different pools of documents. It should be noted that there is no "true" relative size among these four pools of documents as they are all determined by their indexing policy.

	EconLit	ISI citation databases	Top 20 journals	Google Scholar
Mathematical and quantitative methods	36	38	21	50
Economic history	17	17	13	19
Schools of economic thought and methodology	19	15	07	13
Health economics	28	30	60	18

Table 5. Relative sizes of four economics specialties in per cent: journal articles published in 2005.

The table shows that it is evident that these four pools of documents are not duplicating the same picture of productivity in these four economics specialties. Turning to the implications of the uneven coverage of research traditions, table 6 provides an overview of the relative sizes of the three research traditions.

	PsycINFO	ISI citation databases	Top 20 journals	Google Scholar
Cognitive therapy	31	36	69	41
Behavior therapy & behaviour modification	14	20	28	22
Psychoanalytical therapy	54	44	03	37

Table 6. Relative sizes of three psychological research traditions in per cent: journal articles published in 2005.

Like the specialties within economics, the relative sizes of the three research traditions also vary considerably using different pools of documents. The problems of uneven coverage of specialties within economics are therefore also evident when it comes to research traditions in psychology. The four pools of documents are not replicating the same picture of productivity in the three Psychological research traditions. The implications for bibliometric research are discussed, and precautions which need to be taken are outlined.

4.2. The integration of open access journals in the scholarly communication system: Three science fields

The greatest number of OA journals is found in the sciences and their influence is growing. However, the influence of OAJs is not growing with the same rate in all fields as field differences are indicated in the existing literature in terms of the number and the acceptance of OA journals. The sciences are undeniably leaders in establishing OA journals; however, they are distributed unevenly within the sciences (Borgman, 2007: 186).

There are only a few studies on the acceptance and thereby integration of these OA journals in the scholarly communication system. Even fewer studies provide insight into the differences across disciplines.

The article analyses a potential keenness or reluctance of authors (or editors) to integrate OA journals in the reference lists of the accepted publications in non-OA journals and OA journals. The key issue is not the accessibility of single publications but on the perception of quality tied to OA journals in various disciplines. Thus, self-archived non-OAJ articles do not distort the point made here. However, the vast field differences in use of OA and choice of OA model do necessitate that the fields are analysed separately. This article presents an analysis of the citing behaviour in journals within three science fields: biology, mathematics, and pharmacy and pharmacology. They were selected on the basis of the number of OA journals which varies considerably among disciplines.

The methods consist of a statistical analysis of OA journals as well as non-OA journals including both the citing and cited side of the journal to journal citations. The approach is similar to the one used by Baldi (1997, 1998) on document level. Using multiple linear regressions on both cited and citing journals enables controlling for different characteristics of the journals as well as for their degree of interaction or dependency. The variables in the linear regressions consisted of a dependent variable and a number of independent variables. The dependent variable is the relative dependency of the citing journal on the cited journal. The relative dependency of journals is described by the following independent variables: sub-discipline, JIF, publication patterns, OA and variables describing the relationship of the citing and cited journal.

The multivariate linear regression confirms the many similarities between OA journals and non-OA journals which are in accordance with the resemblance in function. However, the results also point to dissimilarities. Causation cannot be determined in the present study; however, it is clear that in some fields authors publishing in OA journals are demonstrating different citing behaviour than authors publishing in non-OA journals. Table 7 summarizes the findings with respect to citation behaviour towards OA journals for the three analysed disciplines. As was already apparent in the presentation of the statistical analyses above there are great field differences in the integration of OA journals.

Table 7. Summary of OA inclusion in three sciences and subdivisions hereof.

	Overall	**Non-OA**	**OA**
Biology	0	0	+
Mathematics	(+)	+	0
Pharmacy and pharmacology	-	-	0

+ indicates OA inclusion, - indicates exclusion while 0 indicates neutrality. Sign shown in parentheses indicates significance at the .1 significance level.

Within biology the non-OA journals are OA neutral with respect to their citing behaviour towards other OA journals. The OA journals within biology are OA including, giving more citations to OA journals in the sample than to other similar journals. Within pharmacy and pharmacology the non-OA journals as well as the OA journals are OA excluding, giving less citations to OA journals in the sample than to other similar journals. Finally, within mathematics non-OA journals are OA including whereas OA journals are neutral. Consequently, even within OA journals there is no guarantee of acceptance and integration of OA journals in general on the level we would expect based on a comparison with non-OA journals with similar characteristics.

4.3. Attracted to open access journals: A bibliometric author analysis in the field of biology

Scholars from developing countries have limited access to research publications due to expensive subscription costs. However, the open access movement is challenging the constraint to access. One of the implications of viewing authors from developing countries as great beneficiaries of open access is that those authors are more likely to perceive OA positively than authors from developed countries as it enables them to access research that they otherwise would have had limited access to. Correspondingly, authors from developed countries should perceive OA less positively than authors from developing countries as they already have access to the necessary research publications within their field.

The article presents an analysis of the use of open access by researchers from developing countries and is thus a supplement to the existing author surveys and interviews. The aim of the article is to examine whether the behaviour of authors can be tied to author characteristics (in terms of geographic location) or journal characteristics (being familiar with the open access principle). The main research question can be elaborated into the following sub-questions (1) Are authors from developing countries more attracted to publishing in OA journals? Are OA journals thus characterised by a greater share of authors from developing countries than traditional subscription based journals? (2) Do authors from developed and developing countries cite differently in OA journals than authors of same nationalities publishing in NOA journals?

The methods applied in the study are bibliometric analyses of both publishing behaviour and citing behaviour in relations to OA publishing to provide evidence of the impact of open access on developing countries. The data is based on two analyses: the first is a publication analysis and the second is a citation analysis, however, the starting points of the two analyses are the same. The dataset collected for the analyses includes journals from the discipline of biology using the classification scheme of Ulrich's Periodicals Directory™ (Ulrich's). Biology is a field characterised by a large number of OA journals making it a suitable field for this study. The dataset consists of a selection of 150 journals that are either general biology journals or belong to one or several of the following sub-disciplines: entomology, zoology, microbiology, biochemistry, genetics, and biotechnology.

The results of the multivariate linear regression show that open access journals are not characterised by a different composition of authors than the traditional toll access journals. OA journals do not differ from non-OA journals with respect to the share of publications written by authors from developing countries. The difference in the percentage of authors from developing countries in non-OA and OA journals is consequently not related to the OA status of the journal but (at least to some extent) other variables.

Furthermore, the results show that authors from developing countries do not cite open access more than authors from developed countries. However, OA journals can be characterised by attracting a certain group of authors as the results show that although authors from developing and developed countries do not differ in terms of citing OA

journals, publications co-authored by authors from developed countries and authors from developing countries differ from the two former groups. It is argued in the article that authors from developing countries are not more attracted to open access journals than authors from developed countries. Further research to be done in this area is suggested.

4.4. The effects of open access on un-published documents: A case study of economics working papers

A perceived major benefit for authors making their publications available OA is the higher number of citations to OA publications presumably due to higher visibility and /or access. The existing analyses of the open access advantage have mainly focused on comparing citation data for archived and not archived journal articles (e.g. Metcalfe, 2006; Moed, 2007; Norris, Oppenheim & Rowland, 2008) or OA and non-OA articles in hybrid journals (Eysenbach, 2006; Davis, 2008). Little attention has been paid to comparing citation data for publications in earlier stages than the journal article.

The open access advantage of publications in earlier stages than the formal publication is particularly interesting as such publications have had limited visibility before the Internet enabled wider dissemination. The Internet has made it possible to disseminate the earlier versions of publications worldwide and research is made available on e.g. personal and institutional websites. Consequently, there has been a growth in both the number (Meadows, 1998) and forms (Farace, 1997). The aim of this study is to investigate if un-published economics literature increases in numbers, visibility and consequently citations over a ten-year period.

The article examines the role of working papers in economics during a ten-year period (1996 to 2005). The article includes two separate analyses. First, the share of grey literature of the publications is determined by calculating the percentage of publications in field specific databases made up by working papers. That analysis provides evidence of visibility and does consequently not provide evidence for the actual numbers and shares of working papers. Secondly, a sample of economics working paper series from 1996 to 2005 (about 2000 working papers in total) was analysed to detect a possible increase in citations to working papers.

It shows that working papers are increasingly becoming visible in the field specific databases. The impact of working papers is relatively low; however, high impact working paper series have citation rate levels similar to the low impact journals in the field. There is no tendency to an increase in impact during the ten years which is the case for the high impact journals.

The results of this study relate to the literature on credit assigning and open access advantage. The results indicate that there is a relatively modest difference in citation rates of low impact journals and high impact working paper series. The difference between the two in citation rates seems to be relatively stable throughout the ten-year period, although with some fluctuations. The clear-cut boundary between published and unpublished publications as well as refereed and un-refereed publications seems less clear-cut as measured by citation rates. Field specific weighting schemes could be worth considering as means to capture differences across fields in the recognition of specific publication types. Economics appears to be a field with relatively high recognition of working paper even though it is typically un-refereed and a form of grey literature. Secondly, the results relate to the open access advantage. The article does not claim to determine causation but rather point to potential developments over time in the citation rates of open access publications that have not previously been focus of attention. The development over time does not indicate that the working paper series have received more citations due to the transformation from print to online version during the years 1998 to 2000. Consequently, the result of this study does not provide evidence of an open access advantage for working papers in economics.

4.5. Turning working papers into articles: An exercise in micro-bibliometrics

This article focuses on the process of scientific and scholarly communication. Data of open access publications on the Internet does not only provide a supplement to the traditional citation indexes. It also enables analysis of the microprocesses and daily practices which constitute scientific communication.

The article focuses on the process of writing that precedes the publication of formal research articles in the scientific and scholarly literature. In this respect, the use of the Web by researchers has basically opened up the area of what formerly was called "the grey literature" (Mili, 2000). Whereas working papers, pre-prints and other forms of grey literature were sometimes difficult to acquire in the past, they are now increasingly available on the Web, both in dedicated repositories and on the personal home pages of researchers (Kling, 2004). The article address the following question: how do authors reformat their working papers in order to turn them into a formal journal article?

The link between a journal article and a preceding version of the publication can be established relatively straightforwardly within those disciplines that are using un-refereed early versions of publications as means to communicate research. Some disciplines have well established traditions of paper manuscript publication going back to the 1960s and economics is one of these disciplines. The analyses are based on a sample of economics working papers subsequently published as journal articles. Binomial logistic regression models are used to analyse precise mechanisms at work in the transformation of working papers into journal articles in the field of economics. The data is analysed in two separate models and in both analyses the dependent variable only had two possible outcomes: 0 and 1. In the first model, 0 represents references that are present in both WP and JA, and 1 represents references that are only present in WP. Correspondingly, in the second model, 0 represents references that are present in both WP and JA and 1 represents references that are only present in the JA. The explanatory variables can be grouped into the following two categories: publication and reference characteristics.

The study unveils a fine-grained process of adapting working papers to their new context as journal article by deleting and adding literature references. The author tries to maximize the fit between the manuscript and the scholarly journal in which she wants the paper published. In this process, deleting and adding references is a very important and telling activity. The probability that a reference will be deleted is influenced by characteristics of both the working paper and the reference. The process of adapting working papers is argued to be best described as "sculpting" as this both reflects that the core structure of the working paper is maintained while important changes are made to several of its components.

To sum up, the article proposes that a better understanding of the precise mechanisms at work in the transformation of working papers into journal articles is highly relevant in a scholarly universe where open access to working papers seems to increasingly shape and transform the reading habits of working scientists and scholars, perhaps already more so than access to the formal journal literature. The article includes a call for more research in line with the micro-bibliometrics and the modelling approach applied in the article.

5. Conclusions

This section summarises the research findings in relation to the research questions. Following this, the results of the articles in relation to the theoretical framework presented in section 3 are analysed. Finally, implications and future investigation are discussed.

The dissertation addresses the overall research question: What are the effects of open access on scholarly communication? The question is been addressed on three different levels. The two general studies focus on introductory studies of open access as an element in scholarly communication and the first study finds that using an open access based database does affect the visibility of research traditions and specialties in relation to existing databases with an available and consistent indexing policy. However, the study also reveals that there is no consistency among traditional databases either in the degree of visibility on an intra-disciplinary level. The second study focuses on the assumption that open access is a pronounced science phenomenon and shows that full OA coverage is not just a matter of time as there are great differences in the integration of open access journals in the sciences that are generally supposed to be the leaders in establishing and incorporating OA journals.

There are two specific studies focusing on two of the most well-known tales of OA. The third study looks into the assumption that OA is of greatest benefit to the developing countries and finds that OA does not significantly affect authors from the developing countries in terms of citing OA available research and publishing their own research in OA journals. The fourth study deals with the open access citation advantage and finds that there are no implications of open access measured by citations for working papers which have received little attention before the Internet.

Finally, the perspectival study opens up new perspectives within the field of OA as it shows that open access enables studies of a publication and its references at different stages in the process.

Before summing up and answering the overall research question it must be noted that it is complex to identify the new forms of social order emerging due to the Internet and thereby separating them from new social phenomena that are not sociologically new. Following Barjak (2006) there are two major concerns: (1) causation and (2) distinguishing between function and technology. First, determining causation is a challenge and the problem of causation is well-known in the social sciences. A study that illustrates association is much simpler because in order to determine causation a study must control for the effect of other variables. Second, in the case of open access function cannot be established on a general level. OA journals and traditional toll-access journals fulfil the same functions; however, in other cases determining functions may be a grey area. An example of a grey area for determining function is lab notes. Electronic notebooks, like the paper versions, record the daily research, experiments and procedures performed in a laboratory. A lab notebook is often considered to be a legal document and may be used in a court of law as evidence (Nature editorial, 2007). Lab notebooks are thus considered to be private property and researchers fear that open notebook science may lead to others stealing their ideas (Bacon, 2008). Lab notebooks made available on the Internet may thus not necessarily reflect the content of the lab notebook with restricted access. Consequently, it can be argued that the functions fulfilled by the open lab notebooks are not the same as those of the traditional lab notebooks with restricted access. Finally, open access cannot be characterised exclusively as a technology as the technology is used in other contexts as well. An example is the technology used for producing OA journals that is also used for producing online versions of print journals.

Bearing these limitations in mind, the overall research question can be answered as follows: on the basis of the attached articles, open access does not appear to change the functions of scholarly communication; however, open access as means to communicate research is preferred more in some disciplines, specialties and research traditions. Finally, open access provides new opportunities for science studies as it allows us to analyse different stages of a scholar's work with a publication.

5.1. Research findings in light of the theoretical framework

In the following the findings of the articles will be seen in the light of the theoretical framework. The structure of the following will follow that of the theoretical framework.

5.1.1. Citation theories

Citation theories have been tested empirically since the 1980s (Nicolaisen, 2007: 618). Nicolaisen (2007) provide an overview of tests of the normative theory and tests of the persuasion hypothesis. Bornmann and Daniel (2006: 50) use a different approach in structuring the empirical studies that "offer motivations for, or categories of, citations and their use". They divide into two approaches: "(1) context or content analyses; and (2) postal surveys or face-to-face interviews of scientists on the topic of citing behaviour" (Bornmann & Daniel, 2006: 50).

Neither of the attached articles are tests of one or another citation theory, however, the results of some of the articles have implications for citation theories. They lend support to one or more citation theories and challenge others. It should be noted that citation theories are not necessarily seen as competing. Bornmann and Daniel (2006: 70) argue:

> Cole (1992) distinguishes between local knowledge outcomes and communal knowledge outcomes. A local knowledge outcome is scientific work produced in a particular context by one or more scientist and may be influenced by social processes. A communal knowledge outcome is work that is accepted by the relevant scientific community as important and correct (the core of research), and it is more or less uninfluenced by social variables and processes. According to Cole (1992), therefore, at the micro-level (local knowledge outcome) we can agree with the position of the constructivists that the content of solutions to scientific problems is developed in a social context and through a series of social processes. In this sense, the content of science is socially constructed. At the macro-level (communal knowledge outcome), in phases in which "normal science" is conducted, the normative theory of science is correct. Core knowledge is

> characterized by virtually universal consensus. Scientists accept this knowledge as a given and as a starting point for their research.

However, accepting the constructivist theory on a micro-level and the normative theory on a macro-level leaves us unable to understand the nature of the citation. As Nicolaisen argues that the understanding of the citation builds on an understanding of the reference implying that "we need a theory of citing that explains why authors cite the way they do" (Nicolaisen, 2007). According to Nicolaisen (2004) the normative and constructivist citation theories are characterised by two different views of science and scholarship. He argues that "in order to comprehend their differences and assess their strength and weaknesses, one needs to study their philosophical and sociological foundations" (Nicolaisen, 2004: 32). Consequently, in the following it is assumed that some results may lend support to some theories and thus not to others.

The article entitled "Attracted to open access journals: A bibliometric author analysis in the field of biology" is based on an analysis of publishing and citing behaviour of authors from developed and developing countries. The results of the study indicate that having controlled for other factors authors from developing countries are not publishing more in OA journals than authors from developed countries. Furthermore, the results indicate that authors from developing countries do not cite OA journals more than authors from developed countries. Authors from developed and developing countries thus have similar publishing and citing behaviour. The results therefore indicate that although access is an obstacle for many authors from developing countries it does not lead them to cite what that can get access to. The results thus lender support to the normative theory of citing and the handicap principle as theory of citing.

A recent article in Science does however report an open access advantage in the developing countries of more than twice as large as in the developed countries (Evans & Reimer, 2009). Apart from the difference in defining developing countries (although both based on gross national income Evans and Reimer measure on a scale and not in two groups) there are further differences. They are not doing the analysis on the level of journals but on the level of articles (due to the large dataset they are not able to characterise on article level and do it on volume level). Furthermore, they include many journals published by non-profit scientific societies that according to Davis (2009, February) often "use the subscription model in tandem with a delayed-access model"

leading him to conclude that "[i]f anyone should be claiming victory, it should be them". Consequently, the two studies are not directly comparable.

The article entitled "Attracted to open access journals: A bibliometric author analysis in the field of biology" also presents results that may be more difficult to interpret. The results show that although authors from developing and developed countries do not differ in terms of citing OA journals, publications by both authors from developed and developing countries differ from the two former groups. Their share of references to OA journals is on average more than 1.5 times greater than the equivalent share in publications by authors from developed countries and about 1.8 times greater than the share in publications by authors from developing countries. The article offers possible interpretations of the results and considers limitations. An alternative interpretation in the light of the theoretical framework is offered by the handicap principle. Authors from developing countries may not be able to send a costly signal that they tend to cite the easy accessible more. They want to have their papers accepted in journals dominated by authors, editors and peer reviewers from developed countries and in order to make that happen they adjust their citing behaviour to that of authors from developed countries. Co-authorships including both authors from developed and developing countries can afford to send that signal and they feel less in need to conform to the behaviour of authors from developed countries. This interpretation, however, implies viewing increased citations to open access journals as costly signals because effective signals are costly. However, it is beyond the scope of the article and the present chapter to determine exactly how costly the signal can be characterised to be.

The article entitled "The effects of open access on un-published documents: A case study of economics working papers" provides results more directly applicable in relation with citation theories. The results show that the impact of working papers is relatively low in the field of economics. There is no clear tendency to an increase in impact during the ten years, which is undoubtedly the tendency of the high impact journals. Consequently, the results do not provide evidence of an open access citation advantage for working papers in economics and as a consequence support the normative theory and the handicap principle.

Finally, there is little support for the constructivist theory of citing in the article entitled "Turning working papers into journal articles: An exercise in micro-bibliometrics".

Seglen (1996) argues that that citing relevant work is ensured if he or she is the editor of the journal submitted. Consequently, references are added to please the editor of the journal that the article is submitted to. However, the results of the article show that the probability of a reference being a journal self-citation is no greater for an added reference than for a reference present in both publications (although it should be noted that there is a slight decrease in the probability of being omitted). MacRoberts and MacRoberts (1996: 440) argue that "citations are added: the names of recognized and respected individuals are prominently displayed at traditional places to persuade an audience". However, the results show that added references are not characterised by being written by authors receiving 500 or more citation or published in core journals (on the contrary) indicating that authors do not add references to "present as authoritative an argument as possible" (MacRoberts & MacRoberts, 1996: 441). There are also results in the article not questioning the constructivist theory of citing. An unpublished document has a more than 25 percentage points higher probability of being omitted than a published article with the same background characteristics. An unpublished document authored by one or more of the same authors has a more than 15 percentage points higher probability of being excluded. MacRoberts and MacRoberts (1996: 436) stress that informal influences are not cited and consequently the formal literature is only the tip of the iceberg (with reference to Edge, 1979). Hence although not challenging the constructivist theory of citing the results do not seriously challenge the normative theory of citing either. Zuckerman and Merton (1971, 1973) acknowledge the limitations of the peer review system but argue (1971, 1973: 494):

> The more specialized the paper, the fewer there are who can responsibly appraise its worth. But while only a few may be fully competent to assess, many more on the periphery of the subject and in other related fields may find the paper relevant to their work. It is for them that the role of the referee as deputy takes on special importance. When a scientist is working on a problem treated in a published article, he can serve as his own referee. He may, in fact, be better qualified to assess its worth than the official referee who helped usher it into print. It is not so much the fellow specialist as the others making use of published results in fields tangential to their own who particularly depend upon the referee system.

Consequently, it is not surprising that authors are reluctant to cite un-refereed publications. Furthermore, it is a possible explanation that authors cite the un-refereed and un-published document in their working paper and should the reference remain un-refereed and un-published when the working paper later is reshaped into a journal article the author may be even more reluctant to cite the un-refereed publication. The author would have to wonder why the document has not been turned into a refereed and published document. The findings are thus conclusive in terms of lending support to specific citation theories.

Summing up, the articles attached contribute to the theoretical framework of citation theories by suggesting that reducing citing behaviour to a matter of access is over-simplifying the complex processes. Access may be an element worth considering but there is much more to it than access. Secondly, the articles attached contribute by showing that adding references when submitting to a journal cannot simply be explained by social constructivist theory of citing.

5.1.2. Mappings of scholarly communication

Existing literature have already commented on developments relating to mappings of scholarly communication. Garvey and Griffith noted as early as in 1964 that the share of publicly available scientific information grows and correspondingly, the share of information available only to a restricted audience decreases (Garvey & Griffith, 1964). Borgman (2007) argues that since the days of print and post the balance has changed between public and private communication. Communication that used to be private (e.g. oral discussions) are now being conducted in public (e.g. online discussion lists, presentations available as slides). Privately circulated publications (e.g. manuscripts, preprints) are now publicly available online. Furthermore, she concludes (2007: 49) that "[o]nline communication has accelerated the amount of informal communication among scholars and simplified the dissemination of formal products of scholarship".

Three of the articles attached relate to mappings of scholarly communication. The article entitled "The integration of open access journals in the scholarly communication system: Three science fields" confirms the considerable differences across disciplines also noted by Garvey, Griffith and collaborators. The findings of the article indicate

discipline specific differences in the use and integration of OA and non-OA journals. To a wide extent the two groups of journals can be described by the same elements. However, the results of the analyses indicate that the two groups of journals differ in terms of the use of OA journals depending on the discipline. In some fields authors publishing in OA journals are demonstrating different citing behaviour than authors publishing in non-OA journals. Consequently, mappings of scholarly communication cannot be done as universal models and must be done on the level of disciplines as done by Garvey, Griffith and collaborators.

The article entitled "Intra-disciplinary differences in database coverage and the consequences for bibliometric research" also contributes to the understanding of mappings of scholarly behaviour. The article points to intra-disciplinary differences as the article documents significant differences in the disciplines of economics and psychology, and reveals quite uneven coverage of economic specialties and psychological research traditions. Some of the differences reported relate to the communicative behaviour within the specialties and research traditions. The extent of the use of open access (both self-archived and OA journals) affects the relative sizes in databases based on open access. The picture depicted in databases based on open access differs considerably from the picture depicted using other pools of documents (showing no consistency either) implying that probably intra-disciplinary characteristics should be considered when mapping scholarly communication. An argument also supported by the article entitled "The effects of open access on un-published documents: A case study of economics working papers" making two relevant points to mappings of scholarly communication. The first is the great intra-disciplinary differences in the use of working papers. In one database working papers ranged from 3 to 38 per cent of the publications and in another from 67 to 97 per cent. The second point being that some of the subject areas with the highest percentages of working papers in one database can be one of the lowest in another.

Finally, the article entitled "Turning working papers into journal articles: An exercise in micro-bibliometrics" takes a starting point in the redundancy of scholarly communication. The study unveils a fine-grained process of adapting working papers to their new context as journal article by deleting and adding literature references. The term suggested to describe the process is "sculpting" as this both reflects that the core structure of the working paper is maintained while important changes are made to

several of its components. Consequently, the article contributes with insight into the redundancy as the content may be the same but the reshaping of that content has a great impact on the reference list.

Summing up, the articles attached contribute to the theoretical framework of mappings of scholarly communication by (a) stressing that the use and integration of OA journals in the existing system of journals is also discipline specific and not a universal phenomenon in all disciplines, (b) suggesting that mappings may benefit from being done on an even more fine grained level than that of disciplines, and finally (c) indicating that although redundancy is a characteristic of scholarly communication, considerable reshaping takes place.

5.2. Research implications and future investigation

The PhD project contributes to the understanding of the effect of open access on scholarly communication. The five bibliometric studies attached investigate different aspects of the overall research question on three different levels.

First of all, the PhD project suggests that open access is a financing model for scholarly communication and not a new form of social order emerging. However, open access is an outstanding opportunity to extend the existing science studies to include the work of scientists at earlier stages than is typically the case presently.

Secondly, the PhD project also contributes theoretically. The contribution to mappings of scholarly communication consists of stressing that OA is discipline specific, by suggesting that mappings should be done on the level of specialties and research traditions and finally by indicating that redundancy in scholarly communication can be characterised by considerable changes from one stage to another. The contribution to citation theories consists of suggesting that there is more to citing behaviour than access and by showing that reshaping a publication cannot simply be explained by social constructivist theory of citing.

Following the results and contributions of the project further research within this area could be done. First of all, the dissertation relies on bibliometric methods, and a number of other methods could be used to study the effects of open access. Recent examples are presented in the following. Bohlin (2004) employs a method stemming from the field of sociology of technology as he uses the Social Construction of Technology (SCOT) methodology to analyse the transformation in scholarly communication. Another alternative is to conduct surveys and numerous surveys investigating various aspects of open access already exist. Examples of surveys including several disciplines are Rowlands, Nicholas, and Huntingdon (2004a, 2004b), Nicholas, Huntingdon and Rowlands (2005) and Swan and Brown (2005). The broad and general surveys tend to suffer from low response rates (these all have response rates of about 5 percent). There are also surveys with a restricted number of journals or faculties as starting point (e.g. Cozzarelli, Fulton & Sullenberger, 2004; Pelizzari, 2003; Richardson & Saxby, 2004; Schroter & Tite, 2006). Generalisability is obviously an issue, but on the other hand response rates tend to be higher. Finally, an alternative method for analysing the effects of open access is to perform interviews (e.g. Park & Qin, 2007; Pickton & McKnight, 2006; Schroter, Tite and Smith, 2005). The results can hardly be generalised but may offer more explanations or interpretations.

There is no self-evident choice of method, and probably a variety of methods is preferable. The triangulation of methods can be done comparing several unique studies or in studies combining several methods as done in the following examples. Zuccala, Oppenheim and Dhiensa (2008) do use different methods to investigate the same research question; however, on two different samples. They use face-to-face interviews of managers of five different types of repositories and a Web-based survey of users. Wilson and Tenopir (2008) use local citation analysis and survey of journal use and reading patterns for evaluating the research collection of an academic library. The methods provided both confirmatory and contradictory results and they found the mix of methods tremendously useful in evaluating library research collections. Consequently, future research using bibliometrics as well as other methods are valuable contributions to the field.

The dissertation has restricted scholarly communication to communication among scientists and future research including lay persons or the public in general could be valuable as their access situation is completely different from that of scholars (Zuccala,

2009). The effect of open access for lay persons would be extremely difficult to capture bibliometrically and other methods would be more appropriate. Furthermore, the dissertation has focused on scholarly publications defined to exclude publications not containing research but e.g. discussing research or delivering news about research. Future research on the changing scholarly communication may include e.g. blogs, RSS feeds and discussion lists. Recent studies investigate aspects of these communication forms. Luzón (2009) analyses the function of links in a corpus of academic blogs and finds that links are strategically used by academic bloggers for several purposes. Plotin (2009) undertake an analysis of the scholarly culture within law to explain why scholarly legal communication has taken the direction it has in the digital age. The use of legal blogs is included in the study.

Finally, open access is to great extent a question of financing models. Consequently, more work should be done considering various financing models and their consequences for libraries, scholars, universities, publishers. There are recent examples of studies analysing various financing models. The study by Hagenhoff, Blumenstiel and Ortelbach (2008) analyses the consequences across disciplines of the author-pays business model based on empirical data. Houghton et al. (2009) examine the costs and benefits of three alternative models for scholarly publishing (i.e. subscription publishing, open access publishing and self-archiving). Future work could also benefit from a greater focus on lay persons and developing countries with respect to specific financing models.

6. Literature

Afzal, W.; Roland, D. and Al-Squri, M. N. (2009). Information asymmetry and product valuation: an exploratory study Journal of Information Science, 35(2), 192-203.

Albert, K. M. (2006). Open access: implications for scholarly publishing and medical libraries. Journal of the Medical Library Association, 94(3), 253–262.

American Psychological Association (1963). A comparison of scientific information exchange activities at three levels of psychological meetings. APA-PSIEP Report #8. Project on scientific Information Exchange in Psychology. Washington, D.C., American Psychological Association.

American Psychological Association (1963). A general study of the annual convention of the American Psychological Association. APA-PSIEP Report #3. Project on scientific Information Exchange in Psychology. Washington, D.C., American Psychological Association.

American Psychological Association (1963). An informal study of the preparations of chapters for the Annual Review of Psychology. APA-PSIEP #2. Project on scientific Information Exchange in Psychology. Washington, D.C., American Psychological Association.

American Psychological Association (1963). Archival journal articles: Their authors and the processes involved in their production. APA-PSIEP Report #7. Project on scientific Information Exchange in Psychology. Washington, D.C., American Psychological Association.

American Psychological Association (1963). Convention attendants and their use of the convention as a source of scientific information. (August). APA-PSIEP Report #4. Project on scientific Information Exchange in Psychology. Washington, D.C., American Psychological Association.

American Psychological Association (1963). Convention participants and the dissemination of information at scientific meetings. APA-PSIEP Report #5. Project on scientific Information Exchange in Psychology. Washington, D.C., American Psychological Association.

American Psychological Association (1963). Scientific activity and information problems of selected psychologists: A preliminary survey. APA–PSIEP Report #1. Project on scientific Information Exchange in Psychology. Washington, D.C., American Psychological Association.

American Psychological Association (1963). The Publication Fate of Formal Presentations at the 1957 convention of the American Psychological Association. APA-PSIEP Report #6. Project on scientific Information Exchange in Psychology. Washington, D.C., American Psychological Association.

American Psychological Association (1963). The use of scientific journals by psychologists and the readership of current journal articles. APA-PSIEP Report #9. Project on scientific Information Exchange in Psychology. Washington, D.C., American Psychological Association.

American Psychological Association (1964). A preliminary study of information exchange activities of foreign psychologists and a comparison of such activities with those occurring in the United States. APA-PSIEP Report #10. Project on scientific Information Exchange in Psychology. Washington, D.C., American Psychological Association.

American Psychological Association (1964). The discovery and dissemination of scientific information among psychologists in two research environments. APA-PSIEP Report #11. Project on scientific Information Exchange in Psychology. Washington, D.C., American Psychological Association.

American Psychological Association (1965). The role of the technical report in the dissemination of scientific information. APA-PSIEP Report #13. Project on scientific Information Exchange in Psychology. Washington, D.C., American Psychological Association.

American Psychological Association (1965). The use of books as a medium for the dissemination of scientific information. APA-PSIEP Report #14. Project on scientific Information Exchange in Psychology. Washington, D.C., American Psychological Association.

American Psychological Association (1965). Theoretical and methodological considerations of undertaking innovations in scientific information exchange. APA-PSIEP Report #12. Project on scientific Information Exchange in Psychology. Washington, D.C., American Psychological Association.

American Psychological Association (1966). A study of *Psychological Abstracts*: Some findings on its current functions and operation and a proposed plan for innovation. APA-PSIEP Report #15. Project on scientific Information Exchange in Psychology. Washington, D.C., American Psychological Association.

American Psychological Association (1966). Innovations in scientific communication in psychology. APA-PSIEP Report #16. Project on scientific Information Exchange in Psychology. Washington, D.C., American Psychological Association.

American Psychological Association (1967). The use of scientific information in the undergraduate teaching of psychology. APA-PSIEP Report #17. Project on scientific Information Exchange in Psychology. Washington, D.C., American Psychological Association.

American Psychological Association (1968). Information exchange at the 1966 American Psychological Association Annual Convention and the function of the convention. Proceedings in such exchange. APA-PSIEP Report #18. Project on scientific Information Exchange in Psychology. Washington, D.C., American Psychological Association.

American Psychological Association (1968). Information exchange activities involved in psychological work. APA-PSIEP Report #19. Project on scientific Information Exchange in Psychology. Washington, D.C., American Psychological Association.

American Psychological Association (1968). Networks of information communication among scientifically productive psychologists: An exploratory study. APA-PSIEP Report #21. Project on scientific Information Exchange in Psychology. Washington, D.C., American Psychological Association.

American Psychological Association (1968). Scientific communication at the XVIII International Congress of Psychology, Moscow, 1966 and its implications for the design and operation of international meetings. Supplement to Volume II, Reports of the American Psychological Association's project on scientific information exchange in psychology, APA-PSIEP #20. Project on scientific Information Exchange in Psychology. Washington, D.C., American Psychological Association.

Antelman, K. (2004). Do open-access articles have a greater research impact? College and Research Libraries, 65(5), 372-382.

Azar, O.H. (2007). The Slowdown in First-Response Times of Economics Journals: Can it Be Beneficial? Economic Inquiry, 45(1), 179-187.

Bacon, D. (2008). Pseudo Open Notebook Science? The Quantum Pontiff Science Blog. Available: http://scienceblogs.com/pontiff/2008/06/pseudo_open_notebook_science.php. Visited April, 2009.

Bakkalbasi, N.; Bauer, K.; Glover, J., and Wang, L.. (2006). Three options for citation tracking: Google Scholar, Scopus and Web of Science. Available: http://eprints.rclis.org/archive/00006080/. Visited April, 2009.

Baldi, S. (1997). A network approach to the analysis of citation flows: A comparative study of two research areas in the natural and the social sciences. Department of Sociology, Ohio State University, Colombus.

Baldi, S. (1998). Normative versus social constructivist processes in the allocation of citations: a network analytic model. American Sociological Review, 63(6), 829-846.

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.

Bazerman, C. (1984). Modern evolution of the experimental report in physics: spectroscopic articles in Physical Review, 1893-1980. Social Studies of Science, 14(2), 163-196.

Bazerman, C. (1985). Physicists reading physics: schema-laden purposes and purpose-laden schema. Written Communication 2(1), 3-23.

Ben-David, J. & Sullivan, T. A. (1975). Sociology of Science. Annual Review of Sociology, 1, 203-222.

Björk, B-C., Roos, A. and Lauri, M. (2009). Scientific journal publishing: yearly volume and open access availability. Information Research, 14(1) paper 391. Available: http://InformationR.net/ir/14-1/paper391.html. Visited April, 2009.

Bloor, D. (1976). Knowledge and social imagery. London: Routledge & Kegan Paul.

Bohlin, I. (2004). Communication regimes in competition: the current transition in scholarly communication seen through the lens of the sociology of technology. Social Studies of Science, 34(3), 365-391.

Bollen, J. and Van de Sompel, H. (2008). Usage Impact Factor: the effects of sample characteristics on usage-based impact metrics. Journal of the American Society for Information Science and Technology, 59(1), 136-149.

Borgman, C. L. (1990). Editor's introduction. In C. L. Borgman (Ed.), Scholarly Communication and Bibliometrics, (pp. 10-27). Newbury Park, CA: Sage Publications.

Borgman, C. L. (2007). Scholarship in the digital age: information, infrastructure, and the Internet. Cambridge, MA: MIT Press.

Borgman, C. L. (2000). Digital libraries and the continuum of scholarly communication. Journal of Documentation, 56(4), 412-430.

Borgman, C. L. and Furner, J. (2002). Scholarly Communication and Bibliometrics. In B. Cronin (Ed.), Annual Review of Information Science and Technology, 36. 3-72.

Bornmann, L. & Daniel, L-D., 2008. What do citation counts measure? A review of studies on citing behavior. Journal of Documentation. 64(1), 45-80.

Brody, T. and Harnad, S. (2005). Keynote Lecture: Providing Open Access to peer-reviewed articles to maximise and measure their research impact. Available: http://www.oai.unizh.ch/symposium/docs/Harnad.ppt. Visited April, 2009.

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, 57(8), 1060-1072.

Bronstein, J. and Baruchson-Arbib, S. (2008). The application of cost—benefit and least effort theories in studies of information seeking behavior of humanities scholars: the case of Jewish studies scholars in Israel. Journal of Information Science, 34(2), 131-144.

Callon, M. (1995). Four Models for the Dynamics of Science. In Handbook of Science and Technology Studies, edited by S. Jasanoff, G. E. Markle, J. C. Petersen and T. Pinch. Thousand Oaks, CA: Sage.

Chan, L. (2008). Comment on blog post entitled: Open Access: No Benefit for Poor Scientists. http://scholarlykitchen.sspnet.org/2009/01/14/oa-developing-nations/. Visited April, 2009.

Chan, L. and Kirsop, B. (2001). Open Archiving Opportunities for Developing Countries: towards equitable distributionof global knowledge. Ariadne, 30. Available: http://www.ariadne.ac.uk/issue30/oai-chan/intro.html. Visited April, 2009.

Chen, C. and Lobo, N. (2006). Analyzing and visualizing the dynamics of scientific frontiers and knowledge diffusion. in Ghaoui, C. ed. Encyclopedia of Human-Computer Interaction, Idea Group Reference, 24-30.

Cole, S. (1992), Making Science. Between Nature and Society, Harvard University Press,
Cambridge, MA.

Cole, J.R. (2000). A short history of the use of citations as a measure of the impact of scientific and scholarly work. In: B. Cronin and H.B. Atkins, Editors, The web of knowledge. A Festschrift in honor of Eugene Garfield, Information Today, Medford, NJ, USA (2000), 281–300.

Cole, S. and Cole, J. R. (1981). Chance and Consensus in Peer Review. Science, 214, 881-86.

Collin, F. (1997). Social reality. London: Routledge.

Collins, H. M. (1981). Stages in the empirical programme of relativism. Social Studies of Science, 11(1), 3-10.

Cozzarelli, N. R., Fulton, K. R. and Sullenberger, D. M. (2004), "Results of a PNAS author survey on an open access option for publication", Proceedings of the National Academy of Sciences, 101, 1111, available at: www.pnas.org/cgi/doi/10.1073/pnas.0307315101. Visited April, 2009.

Craig, I. D.; Plume, A. M.; McVeigh, M. E.; Pringle, J. and Amin, M. (2007). Do open access articles have greater citation impact? A critical review of the literature. Journal of Informetrics, 1(3), 239-248. Available: http://dx.doi.org/10.1016/j.joi.2007.04.001. Visited April, 2009.

Crane, D. (1972). Invisible colleges. Chicago: University of Chicago Press.

Cronin, B. (2001). Bibliometrics and beyond: some thoughts on web-based citation analysis. Journal of Information Science, 27(1), 1-7.

Cronin, B. (2003). Scholarly communication and epistemic cultures. Keynote address. Scholarly tribes and tribulations: How tradition and technology are driving disciplinary change. ARL. Washington, DC, October 17, 2003.

Cronin, B., & Atkins, H. B. E. (2000). The Web of Knowledge: A Festschrift in Honor of Eugene Garfield: American Society for Information Science & Technology.

Culnan, M. J. (1985). The dimensions of perceived accessibility to information: Implications for the delivery of information systems and services. Journal of the American Society for Information Science, 36(5), 302-308.

Davis, P. M. (2009). Reward or persuasion? The battle to define the meaning of a citation. Learned Publishing 22(1), 5-11.

Davis, P. M.; Lewenstein, B.; Simon, D.; Booth, J. and Connolly, M. (2008). Open access publishing, article downloads, and citations: randomised controlled trial. British Medical Journal 337:a568.

Davis, P. M. (2008). Author-choice open access publishing in the biological and medical literature: a citation analysis. Journal of the American Society for Information Science and Technology, 60(1), 3-8.

Davis, P. M. (2009, February). Open Access and Global Participation in Science. Blog post. Available: http://scholarlykitchen.sspnet.org/2009/02/19/oa-global-participation/. Visited April, 2009.

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

de Solla Price, D. (1993). Little science, big science. New York: Columbia Univ. Press.

Dervin, B., and Nilan, M. (1986). Information needs and uses. In M. E. Williams (Ed.), Annual Review of Information Science and Technology, 21, 3-33.

Dewatripont, M.; Ginsburgh, V.; Legros, P.; Walckiers, A.; Devroey, J. P.; Dujardin, M.; Vandooren, F.; Dubois, P.; Foncel, J.; Ivaldi, M. and Heusse, M. D. (2006). Study on the Economic and Technical Evolution of the Scientific Publication Markets in Europe" (online). URL: http://europa.eu.int/comm/research/science-society/pdf/scientific-publication-study_en.pdf. Visited April, 2009.

Dominguez, M. B. (2006). Economics of open access publishing. Serials, 19(1), 52-60.

Duff, A.S. (1997). 'Some post-war models of the information chain', Journal of librarianship and information science, 29(4): 179-187.

Eichorn, P. and Yankauer, A. (1987). Do authors check their references? A survey of accuracy of references in three public health journals. American Journal of Public Health, 77(8), 1011-1012.

Ellison, G. (2002a). The Slowdown of the Economics Publishing Process. Journal of Political Economy, 110(5), 947−993.

Ellison, G. (2002b). Evolving Standards for Academic Publishing: A q-r Theory. Journal of Political Economy, 110(5), 994−1034.

Evans, J. A. and Reimer, J. (2009). Open Access and Global Participation in Science. Science 20 February: 1025

Evans, J. T.; Nadjari, H. I. and Burchell, S. A. (1990). Quotational and reference accuracy in surgical journals: A continuing peer review problem. Journal of the American Medical Association, 263(10), 1353-1354.

Gans, J. S. and Shepherd, G. B. (1994). How are the mighty fallen: rejected articles by leading economists. Journal of Economic Perspectives, 8, 165–179.

Garvey, W. D. and Compton, B. E. (1967). A program of research in scientific information exchange : Orientation, objectives, and results. Social Science Information, 6, 213-236.

Garvey, W. D. and Gottfredson, S. D. (1976). Changing the System: Innovations in the Interactive Social System of Scientific Communication. Information Processing and Management, 12(3), 165-76.

Garvey, W. D. and Gottfredson, S. D. (1977). Scientific communication as an interactive process. International Forum for Information and Documentation, 2(1), 9-16.

Garvey, W. D. and Griffith, B. C. (1964). Scientific information exchange in psychology. Science, 146(3652), 1655–1659.

Garvey, W. D. and Griffith, B. C. (1964). The structure, objectives and findings of a study of scientific information exchange in psychology. American Documentation, 15, 258-267.

Garvey, W. D. and Griffith, B. C. (1967). Scientific communication as a social system. Science, 157(3792), 1011–1015.

Garvey, W. D., and Griffith, B. C. (1971). Scientific communication: Its role in the conduct of research and creation of knowledge. American Psychologist, 26(4), 349–362.

Garvey, W. D., Tomita, K., and Woolf, P. (1974). The dynamic scientific information user. Information Storage and Retrieval, 10, 115-31.

Garvey, W. D.; Lin, N., and Nelson, C. E. (1970). Communication in the physical and the social sciences. Science, 170(3963), 1166–1173.

Garvey, W. D.; Lin, N., and Nelson, C. E. (1971). Scientific communication behavior of social and physical scientists. International Social Science Journal, 23(3), 256–271.

Garvey, W. D. & Griffith, B. C. (1972). Communication and information processing within scientific disciplines: Empirical findings for psychology. Information Storage and Retrieval, 8, 123-126.

Garvey, W.D. (1979). Communication: the essence of science. Oxford: Pergamon Press.

Garvey, W. D., and Griffith, B. C. (1963). An overview of the structure, objectives, and findings of the American Psychological Association's project on scientific information exchange in psychology. Washington, DC: American Psychological Association.

Gaulé, P. and Maystre N. (2008). Getting cited: does open access help? No cemi-workingpaper-2008-007, CEMI Working Papers from Ecole Polytechnique Fédérale de Lausanne, Chaire en Economie et Management de l'Innovation. Available: http://ilemt.epfl.ch/repec/pdf/cemi-workingpaper-2008-007.pdf. Visited April, 2009.

Gilbert, G. N. and Mulkay, M. (1984). Opening Pandora's box: A sociological analysis of scientists' discourse. Cambridge, England: Cambridge university Press.

Griffith, B. C. (1990). Understanding science: Studies of communication and information. In C. L. Borgman (ed.), Scholarly Communication and Bibliometrics (31-45). Newbury Park, CA: Sage.

Hackett, E. J.; Amsterdamska, O.; Lynch, M. and Wajcman, J. (2008). Introduction. In The Handbook of Science and Technology Studies, edited by E. J. Hackett, O. Amsterdamska, M. Lynch and J. Wajcman. Cambridge, MA, USA: MIT Press. 1-7.

Hagenhoff, S.; Blumenstiel, M. and Ortelbach B. (2008). An Empirical Analysis of the Amount of Publication Fees Serials Review, 34(4), 257-266.

Hajjem, C., Gingras, Y., Brody, T., Carr, L. and Harnad, S. (2005). Open Access to Research Increases Citation Impact. Technical Report, Institut des sciences cognitives, Université du Québec à Montréal.

Hajjem, C., Harnad, S. and Gingras, Y. (2005) Ten-Year Cross-Disciplinary Comparison of the Growth of Open Access and How it Increases Research Citation Impact. IEEE Data Engineering Bulletin 28(4), 39-47.

Hall, G. (2008), Digitize this Book! The politics of new media or why we need open access now, University of Minnesota Press.

Hardy, A. P. (1982). The selection of channels when seeking information: cost—benefit vs least effort. Information Processing and Management 18(6), 289-293.

Harnad, S. (2006). Self-archiving should be mandatory. Interview with Steven Harnad. Available: http://www.researchinformation.info/rijunjul06openaccess4.html. Visited April, 2009.

Harnad, S. (2008a). Davis et al's 1-year Study of Self-Selection Bias: No Self-Archiving Control, No OA Effect, No Conclusion Rapid response to Philip M Davis, Bruce V Lewenstein, Daniel H Simon, James G Booth, and Mathew J L Connolly Open access publishing, article downloads, and citations: randomised controlled trial BMJ 2008; 337: a568. http://www.bmj.com/cgi/eletters/337/jul31_1/a568#199775. Visited April, 2009.

Harnad, S. (2008b). Not a Response: A correction on declared interests. Rapid response to Philip M Davis, Bruce V Lewenstein, Daniel H Simon, James G Booth, and Mathew J L Connolly Open access publishing, article downloads, and citations: randomised controlled trial British Medical Journal 2008; 337: a568. http://www.bmj.com/cgi/eletters/337/jul31_1/a568#199895. Visited April, 2009.

Harnad, S. (2008c). Open Access: The Devil's in the Details. Monday, December 29. 2008. Available: http://openaccess.eprints.org/index.php?/archives/499-Open-Access-The-Devils-in-the-Details.html. Visited April, 2009.

Harnad, S. and Brody, T. (2004). Comparing the impact of open access (OA) vs. non-OA articles in the same journals. D-Lib Magazine, 10(6).

Henneken, E. A., Kurtz, M. J., Eichhorn, G., Accomazzi, A., Grant, C., Thompson, D., and Murray, S. S. (2006). Effect of E-printing on Citation Rates in Astronomy and Physics. Journal of Electronic Publishing, 9(2).

Hess, D. J. (1997). Science studies: An advanced introduction. New York: New York University Press.

Houghton, J. W.; Rasmussen, B.; Sheehan, P. J.; Oppenheim, C.; Morris, A.; Creaser, C.; Greenwood, H.; Summers, M. and Gourlay, A. (2009). Economic Implications of Alternative Scholarly Publishing Models: Exploring the Costs and Benefits, Report to The Joint Information Systems Committee (JISC) by Victoria University & Loughborough University.

Hurd, J. M. (1996). Models of scientific communications systems. In: S. Crawford, et al. (Eds). From print to electronic: the transformation of scientific communication. (pp. 9-33) Medford, NJ: Information Today Inc.

Hurd, J. M. (2000). The transformation of scientific communication: A model for 2020. Journal of the American Society for Information Science, 51(14), 1279-1283.

Hurd, J. M. (2005). Scientific communication: New roles and new players. Science and Technology Libraries 25(1-2), 5-22.

Jacso, P. (2004). Google Scholar. Available: http://www.galegroup.com/reference/archive /200412/googlescholar.html.

Jankowski, N. (Ed.) (in press), e-Research: Transformation in Scholarly Practice. Routledge. Available: http://www.slideshare.net/nickjan/jankowski-chapter-1-e-research-introduction-final-single-spaced-small-font-3-december2008. Visited April, 2009.

Jasanoff, S.; Markle, G. E.; Petersen, J. C. and Pinch, T. (eds) (1995). Handbook of Science and Technology Studies. Thousand Oaks, CA, London & New Delhi: Sage/4S.

Jenkins, C.; Probets, S.G.; Oppenheim, C. and 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.

Jones, A. W. (2003). Impact factors of forensic science and toxicology journals: what do the numbers really mean? Forensic Science International, 133(1-2), 1-8.

Kirshop, B. and Chan, L. (2005). Transforming access to research literature for developing countries. Serials Review, 31(4), 246-255.

Klamer, A. and van Dalen, H. P. (2002). Attention and the art of scientific publishing. Journal of Economic Methodology, 9(3), 289-315.

Kling, R. (2004). The internet and unrefereed scholarly publishing. Annual Review of Information Science and Technology, 38, 591-631.

Knorr Cetina, K. D. (1982). Scientific Communities or Transepistemic Arenas of Research. Social Studies of Science 12(1), 101-130.

Knorr Cetina, K. D. (1995). Laboratory Studies: The Cultural Approach to the Studyof Science. In S. Jasanoff, G. E. Markle, J. C. Petersen and T. J. Pinch (Eds.), Handbook of Science, Technology and Society, Los Angeles, Sage.

Knorr-Cetina, K. D. (1981). The manufacture of knowledge: An essay on the constructivist and contextual nature of science. Oxford: Pergamon Press.

Kousha, K. and Thelwall, M. (2006) Google Scholar Citations and Google Web/URL Citations: A Multi-Discipline Exploratory Analysis. In Proceedings International Workshop on Webometrics, Informetrics and Scientometrics & Seventh COLLNET Meeting, Nancy (France). http://eprints.rclis.org/archive/00006416/. Visited April, 2009.

Kuhn, T. S. (1962). The Structure of Scientific Revolutions, 1st. ed., Chicago: Univ. of Chicago Press.

Kurtz, M J.; Eichhorn, G.; Accomazzi, A.; Grant, C.; Demleitner, M.; Henneken, E. and Murray, S. S. (2005). The effect of use and access on citations. Information Processing and Management, 41(6), 1395-1402.

Latour, B. (1987). Science In Action. How to Follow Scientists and Engineers through Society. Harvard University Press, Cambridge Mass.

Latour, B. and Woolgar, S. (1979). Laboratory life: The social construction of scientific facts. Beverly hills, CA: Sage.

Laudan, L. (1977). Progress and its Problems: Toward a Theory of Scientific Growth. Berkeley, CA: University of California Press.

Lawrence, S. (2001) Online or Invisible? Nature 411 (6837): 521. http://www.neci.nec.com/~lawrence/papers/online-nature01/. Visited April, 2009.

Leckie, G. J.; Pettigrew, K. E. and Sylvain, C. (1996). Modelling the information seeking of professionals: a general model derived from research on engineers, health care professionals and lawyers. Library Quarterly 66(2), 161-193.

Lewison, G. and Sullivan, R. The impact of cancer research: how publications influence UK cancer clinical guidelines. British journal of Cancer, 98(12), 1944-1950.

Lewison, G.; Tootell, S.; Roe, P. and Sullivan, R. (2008). How do the media report cancer research? A study of the UK's BBC website. British journal of Cancer, 99(4), 569-576.

Lievrouw, L. A. (1988). Four programs of research in scientific communication. Knowledge in Society, 1(2), summer, 6-22.

Lievrouw, L. A. (1990). Reconciling structure and process in the study of scholarly communication. In: Christine L. Borgman (Ed.), Scholarly Communication and Bibliometrics, pp. 59-69. Newbury Park, CA: Sage.

Liu, Z. (2006). Print vs electronic resources: a study of user perceptions, preferences, and use. Information Processing and Management, 42(2), 583-592.

Lucas, R. and Willinsky, J. (In Press). Open Access to E-Research. In N. Jankowski (Ed.), e-Research: Transformations in Scholarly Practice. London: Routledge.

Luzón, M. J. (2009). Scholarly hyperwriting: The function of links in academic weblogs. Journal of the American Society for Information Science, 60(1), 75-89.

Mackenzie Owen, J. (2002). The new dissemination of knowledge: digital libraries and institutional roles in scholarly publishing. Journal of Economic Methodology, 9(3), 275-288.

Mackenzie Owen, J. S. and Halm, J. van (1989) Innovation in the information chain: the effects of technological development on the provision of scientific and technical information, London: Routledge.

MacRoberts, M. H. and MacRoberts, B. R. (1996). Problems of Citation Analysis. Scientometrics, 36(3), 435-444.

McCabe, M. J. (2002). Journal Pricing and Mergers: A Portfolio Approach. American Economic Review 92, 259–269.

Meadows, A. J. (1998). Communicating Research. San Diego, CA: Academic Press.

Meadows, A. J., ed. (1991). Knowledge and communication: essays on the information chain, London: Library Association Publishing.

Merton, R. K. (1976). Sociological Ambivalence. New York The Free Press, 1976.

Merton, R. K. (2000). On the Garfield input to the sociology of science: A retrospective collage. In B. Cronin and H. Atkins (Eds.). The Web of knowledge: A Festschrift in honor of Eugene Garfield (435-448). Medford, NJ: Information Today.

Merton, R. K. (1942). The Normative Structure of Science. In: R.K. Merton, The Sociology of Science: Theoretical and Empirical Investigations. Chicago, IL: University of Chicago Press, 1973.

Merton, R. K. (1957, 1973). Priorities in Scientific Discovery. In: R.K. Merton, The Sociology of Science: Theoretical and Empirical Investigations. Chicago, IL: University of Chicago Press.

Merton, R. K. (1977) The Sociology of Science: An Episodic Memoir , pp. 3—141 in R.K.

Merton and J. Gaston, The Sociology of Science in Europe . Carbondale: Southern Illinois University Press.

Metcalfe, T. S. (2005). The Rise and Citation Impact of astro-ph in Major Journals. Bulletin of the American Astronomical Society, 37, 555-557.

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.

Moed, H. F. (2005). Citation Analysis in Research Evaluation. Dordrecht, NL: Springer.

Mulkay, M. (1985). The word and the world: Explorations in the form of sociological analysis. London: Allen & Unwin.

Mulkay, M. and Gilbert, G. N. (1982). What is the ultimate question? Some remarks in the defence of the analysis of scientific discourse. Social Studies of Science, 12, 309-319.

Mullins, N. C. (1972). The development of a scientific specialty: The Phage Group and origins of molecular biology. Minerva 10(1), 51-82.

National Science Board. 2008. Science and Engineering Indicators 2008. Two volumes. Arlington, VA: National Science Foundation. Available: http://www.nsf.gov/statistics/seind08/

Nature Editorial (2007). Share your lab notes. Nature 447, 1-2. Available: http://www.nature.com/nature/journal/v447/n7140/full/447001b.html. Visited April, 2009.

Navarro, F. A. (1999). Reference inaccuracy: are articles cited without being read? Journal of Information Science, 25(5), 423-424.

Neuhaus, C. and Daniel, H-D. (2008). Data sources for performing citation analysis: An overview. Journal of Documentation, 64(2), 193-210.

Nicholas, D.; Huntington, P. and Jamali, H. R. (2007). Open access in context: a user study. Journal of Documentation, 63(6), 853-878.

Nicholas, D.; Huntingdon, P. and Rowlands, I. (2005). Open access publishing: the views of some of the world's senior authors. Journal of Documentation, 61(4), 497-519.

Nicolaisen, J. and Frandsen, T. F. (2007). The handicap principle: A new perspective for LIS research. Proceedings of the 6th International Conference on Conceptions of Library and Information Science: 119-130. Co-published in Information Research, 12(4). http://informationr.net/ir/12-4/colis/colis23.html. Visited April, 2009.

Nicolaisen, J. (2004). Social Behavior and Scientific Practice – Missing Pieces of the Citation Puzzle. PhD-thesis. Copenhagen, DK: Royal School of Library and Information Science.

Nicolaisen, J. (2007). Citation analysis. Annual Review of Information Science and Technology, 41: 609-641.

Norris, M.; Oppenheim, C. and Rowland, F. (2008). The citation advantage of open-access articles. Journal of the American Society for Information Science and Technology, 59(12), 1963-1972.

Noruzi, A. (2005). Google Scholar : the new generation of citation indexes. LIBRI 55(4), 170-180. Available: http://eprints.rclis.org/archive/00005595/. Visited April, 2009.

Ortelbach, B.; Schulb, S. and Hagenhoff, S. (2008). Journal Prices Revisited: A Regression Analysis of Prices in the Scholarly Journal Market. Serials Review 34(3).

Pais, A. (1982). Subtle is the lord: the science and the life of Albert Einstein. Oxford: Clarendon Press.

Paisley, W. (1984). Communication in the communication sciences. In Dervin, B. and Voigt, M. J. (Eds.), Progress in Communication Sciences, V, 1-43.

Park, J. H. (2008). The Relationship between Scholarly Communication and Science and Technology Studies (STS). Journal of Scholarly Publishing 39(3), 257-73.

Park, J. H. and Qin, J. (2007). Exploring the willingness of scholars to accept open access: A grounded theory approach. Journal of Scholarly Publishing, 38(2), 55-84.

Pelizzari, E. (2003). Academic staff use, perception and expectations about open-access archives. A survey of social science sector at Brescia University. Available at: http://eprints.rclis.org/archive/00000737/01/Academic_staff_perception_about_Open_archives.htm. Visited April, 2009.

Perneger, T. V. (2004). Relation between online "hit counts" and subsequent citations: prospective study of research papers in the BMJ. British Medical Journal, 329, 546-547.

Pickton, M. and McKnight, C. (2006). Research students and the Loughborough institutional repository. Journal of Librarianship and Information Science, 38(4), 203-219.

Plotin, S. (2009). Legal Scholarship, Electronic Publishing, and Open Access: Transformation or Steadfast Stagnation? Law Library Journal, 101(1), 31-57.

Regazzi, J. J. and Caliguiri, N. A. (2006). Publisher and author partnership: a changing landscape. Learned Publishing, 19(3), 183-193.

Restivo, S. and Loughlin, J. (1987). Critical sociology of science and scientific validity. Knowledge, 8, 486-508.

Richardson, M. and Saxby, C. (2004). Experimenting with open access publishing. Nature. Available: www.nature.com/nature/focus/accessdebate/12.html. Visited April, 2009.

Ross, T. and Ross, M. (2002). Complete Guide to Self Publishing: Everything You Need to Know to Write, Publish, Promote, and Sell Your Own Book (Self-Publishing 4th Edition). Ohio: Writer's Digest Books.

Rowlands, I.; Nicholas, D. and Huntingdon, P. (2004a). Journal publishing: what do authors want? Available at: http://www.nature.com/nature/focus/accessdebate/31.html. Visited April, 2009.

Rowlands, I.; Nicholas, D. and Huntingdon, P. (2004b). Scholarly communication in the digital environment: what do authors want? Findings of an international survey of author opinion: project report. London: Centre for Information Behaviour and Evaluation of Research, Department of Information Science, City University, 2004.

Schroter, S. and Tite, L. (2006). Open Access publishing and author-pays business models: a survey of authors' knowledge and perceptions. Journal of the Royal Society of Medicine, 99, 141-148.

Schroter, S.; Tite, L. and Smith, R. Perceptions of open access publishing: interviews with journal authors. BMJ, Published 26 January 2005.

Seglen, P. O. (1996). Bruk af siteringer og tidsskriftimpaktfaktor til forskningsevaluering. Biblioteksarbejde, 48, 27-34.

Shadbolt, N., Brody, T., Carr, L. and Harnad, S. (2006) The Open Research Web: A Preview of the Optimal and the Inevitable, in Jacobs, N., Eds. Open Access: Key Strategic, Technical and Economic Aspects, chapter 21. Chandos.

Shapin S. 1994. A Social History of Truth: Civility and Science in Seventeenth-Century England. Chicago: Univ. Chicago Press.

Sismondo, S. (1993). Some Social Constructions. Social Studies of Science, 23(3), 515-553.

Small, H. (1997). Belver. C. Giffith and John Irvine and Ben R. Martin (as a team) win the 1997 Derek de Solla Price award. Scientometrics, 40(3): 357-362.

Smelser, N. J. (1988). Social structure. In Handbook of sociology, edited by Neil J. Smelser. Beverly Hills, CA: Sage.

Spence, A. M. (1973). Job Market Signaling. The Quarterly Journal of Economics, 87(3), 355-374.

Spence, A. M. (1974). An economist's view of information. Annual Review of Information Science and Technology, 9, 57-78.

Spence, A. M. (2002). Signaling in Retrospect and the Informational Structure of Markets. American Economic Review 92(3): 434–459.

Spink, A., Robins, D. & Schamber, L. (1998). Use of scholarly book reviews: implications for electronic publishing and scholarly communication. Journal of the American Society for Information Science, 49(4), 364-374.

Suber, P. (2003). Removing the Barriers to Research: An Introduction to Open Access for Librarians. College & Research Libraries News, 64(113), 92-94.

Suber, P. (2003, June). Bethesda Statement on Open Access Publishing. Available: http://www.earlham.edu/~peters/fos/bethesda.htm. Visited April, 2009.

Suber, P. (2004a). Guide to the Open Access Movement. Formerly called the Guide to the Free Online Scholarship Movement. Available: http://www.earlham.edu/~peters/fos/guide.htm#fos Visited April, 2009.

Suber, P. (2004b). Praising progress, preserving precision Welcome to the SPARC Open Access Newsletter, issue #77. September 2, 2004.

Suber, P. (2008a). Open Access Overview. Focusing on open access to peer-reviewed research articles and their preprints. Available: http://www.earlham.edu/~peters/fos/overview.htm Visited April, 2009.

Suber, P. (2008b). Welcome to the SPARC Open Access Newsletter, issue #124 August 2, 2008.

Tague-Sutcliffe, J. (1992). An introduction to informetrics. Information Processing and Management 28(1), 1-3.

Talja, S. and Hartel, J. (2007). Revisiting the user-centred turn in information science research: an intellectual history perspective. Information Research, 12(4), paper colis04. Available: http://InformationR.net/ir/12-4/colis/colis04.html. Visited April, 2009.

UNISIST (1971). Study Report on the feasibility of a World Science Information System, By the United Nations Educational, Scientific and Cultural Organization and the International Council of Scientific Unions. Paris, UNESCO.

Wang, C. (2003). Electronic publishing: Significant landmarks. In M. A. Drake (Ed.), Encyclopedia of library and information science (2nd ed., Vol. 2, 1011-1015). New York: Dekker.

Weinberg, A. M. (1961). Impact of Large-Scale Science on the United States: Big science is here to stay, but we have yet to make the hard financial and educational choices it imposes. Science, 134(3473), 161-164.

Whitley, R. (1991). The organisation and role of journals in economics and other scientific fields. Working Paper 204. Manchester business School.

Whitley, R. (2000). The intellectual and social organization of the sciences. Oxford University Press. 2nd ed.

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

Wilson, C. S. and Tenopir, C. (2008). Local citation analysis, publishing and reading patterns: Using multiple methods to evaluate faculty use of an academic library's

research collection. Journal of the American Society for Information Science, 59(9), 1393-1408.

Wilson, L. (1942). The academic man: a study in the sociology of a profession. New York.

Wouters, P. and de Vries, R. (2004). Formally Citing the Web. Journal of the American Society for Information Science and Technology, Special Issue Webometrics, 55(14), 1250-1260.

Wouters, P.; Vann, K.; Scharnhorst, A.; Ratto, M.; Hellsten, I.; Fry, J. and Beaulieu, A. (2008). Messy Shapes of Knowledge - STS Explores Informatization, New Media, and Academic Work. The Virtual Knowledge Studio. In: The Handbook of Science and Technology Studies, edited by E. J. Hackett, O. Amsterdamska, M. Lynch and J. Wajcman. Cambridge, MA, USA: MIT Press. 319-352.

Wright, M. and Armstrong, J. S. (2008). The Ombudsman: Verification of Citations: Fawlty Towers of Knowledge? Interfaces, 38(2), 125-139.

Yeomans, J. (2006). CERN's Open Access E-print Coverage in 2006 : Three Quarters Full and Counting. High Energy Physics Libraries Webzine, issue 12, March 2006.

Zahavi, A. 2003. Indirect selection and individual selection in sociobiology: My personal views on theories of social behaviour. Animal Behaviour, 65, 859-863. Available: http://www.tau.ac.il/lifesci/departments/zoology/members/zahavi/zahavi.html. Visited April, 2009.

Zahavi, A., 1975. Mate selection – a selection for a handicap. Journal of Theoretical Biology, 53(1), 205-214.

Zahavi, A., and Zahavi, A. (1997). The Handicap Principle. A missing piece of Darwin's puzzle. New-York: Oxford University Press.

Zuccala, A. (2009). The layperson and open access. In B. Cronin (Ed.), Annual Review of Information Science and Technology, 43, 359-396.

Zuccala, A.; Oppenheim, C. and Dhiensa, R. (2008). Managing and evaluating digital repositories. Information Research, 13(1) , 333.

Zuckerman, H. and Merton, R. K. (1971, 1973). Patterns of evaluation in science: institutionalization, Structure and Functions of the referee system. In: R.K. Merton, The Sociology of Science: Theoretical and Empirical Investigations. Chicago, IL: University of Chicago Press.

Zuckerman, H. (1988). The Sociology of Science. In Neil J. Smelser (Ed.), Handbook of Sociology, 511-574. Newbury Park, CA: Sage

7. Appendix

7. 1. Appendix 1. The data material in the APA-PSIEP Reports

Report	**Data material**
Report#1: Scientific activity and information problems of selected psychologists: a preliminary survey.	132 APA [American Psychological Association] members were to keep detailed log of their scientific information activities during a two-week period. 90 agreed to do so and 78 was actually received and used in the study. A more detailed questionnaire was completed by 71 of the 78 participants. Data included information on age and sources of articles read, filing and note-taking, correspondence, use of Psychological Abstracts and difficulties in obtaining information.
Report#2: An Informational Study of the preparation of chapters for the *Annual Review of Psychology.*	81 (of 128) reviewers of Annual Review of Psychology describe their "methods of conducting a literature search on a topic, the difficulties impeding this search, their attitudes and objectives as reviewers, the ways in which responsibilities are divided when there is more than one author of a review, the inadequacies of Psychological Abstracts, and the manner of organizing and presenting the material discovered."
Report#3: A general study of the annual convention of the American Psychological Association.	The study is based on the following data: "programs of the APA conventions for 1936, 1951, 1957 and 1961 were examined, and data collected on various aspects of convention presentations and persons making them". The data consisted of data on e.g. rejection rates,

	number of events, duration of presentations, type of presentation, educational level of participants, and sponsorships.
Report#4: Convention attendants and their use of the convention as a source of scientific information.	On the basis of experience from pretests questionnaires were sent out to every tenth attendant at American Psychological Association convention and every registrant at the Psychonomic Society meetings (both 1962). They received 280 usable (of 409) questionnaires in the former group and 191 (of 242) in the latter group.
Report#5: Convention participants and the dissemination of information at scientific meetings.	Questionnaires were sent to a sample of persons presenting at one of these 1962 scientific meetings: Eastern Psychological Association (EPA), American Psychological Association (APA) and Psychonomic Society (PS). EPA had 323 presentations, 321 questionnaires were sent out and they received 262 usable. At APA there were 1230 presentations, 246 questionnaires were sent out and 189 were usable. PS had 119 presentations, 119 questionnaires were sent out and they received 102 usable.
Report#6: The publication fate of formal presentations at the 1957 convention of the American Psychological Association.	The 987 persons "who actually made a formal presentation of material" at the 1957 convention of the American Psychological Association were sent a questionnaire and 764 were returned. The questionnaire was used to determine the fate of the information presented in formal oral reports at meetings.

Report#7: Archival journal articles: their authors and the processes involved in their production.	Authors of 396 articles were polled "to determine (a) some of the processes, and associated time intervals involved in the writing of an article, and (b) the author's experiences in submitting articles for publication." The 396 articles had 691 authors of which addresses could be found for 668. The 668 resulted three to four months later in 543 usable questionnaires.
Report#8: A comparison of scientific information activities at three levels of psychological meetings.	This study aimed to present the same type of data as report #4 and report#5 "on state associations in order to clarify the role of these meetings in the overall system of dissemination of scientific information in psychology." They polled samples of convention attendants at 11 1963 meetings. They polled 957 out of 972 and received 686 surveys in total. The response rate for attendants at the 11 different meetings ranged from 51 to 82%
Report#9: the use of scientific journals by psychologists and the readership of current journal articles.	The study aimed to (a) investigate the audience for journals and their use of journals and (b) the reading of current journal articles. Questionnaires were sent out to 10% of the members of the American Psychological Association. Of the 2140 persons polled 1187 persons returned usable questionnaires.
Report#10: A preliminary study of information exchange activities of foreign psychologists and a comparison of such activities with those occurring in the United States	A questionnaire was sent to 125 foreign psychologists and they received 93 usable questionnaires. Correspondingly, they sent out a questionnaire to 91 United states psychologists and received 73 usable.

Report#11: The discovery and dissemination of scientific information among psychologists in two research environments.	The first part of the study "deals with those members of a faculty of a large university who were guilding or actively engaged in research in the field of psychology and the media of information exchange (formal and informal, archival and non-archival) which they utilized". Data consisted of tape-recorded interviews and questionnaire of 63 informants. The second part of the study takes place in a government laboratory and data is collected the same way as for the first setting (30 informants).
Report#12: Theoretical and methodological considerations of undertaking innovations in scientific information exchange.	The first part of the study aims to develop "some of the substance and empirical background for a theory of innovation in scientific information exchange". Part 1 does not present empirical findings. The second part "provides a detailed description of the various means of dissemination involved in or related to a convention presentation." Data is gathered using 7 persons that monitored presentations at a convention and questionnaires sent out to three groups (participants, attendants and requestors) identified at these presentations.
Report#13: The role of the technical report in the dissemination of scientific information.	A sample was drawn from the Psychology Section of the 1962 National Register of Scientific and technical Personnel and more than 6000 questionnaires were sent out. 4163 were returned and of these "only 1135 had participated in the writing of at least one technical report issued in 1962" and consequently made out the sample for analysis. The second part of the report is an analysis of

	the 236 persons in the first analysis that had written a report that was later published as a journal article. The study consisted of 116 published versions of 112 technical reports that provided information on the relations between technical report and journal article.
Report#14: The use of books as a medium for the dissemination of scientific information.	Questionnaires were sent out to app. 10% of the membership of American Psychological Association. 1263 of the 2240 psychologists returned usable questionnaires.
Report#15: A study of *Psychological Abstracts*: Some findings on its current functions and operation and a proposed plan for innovation.	A sample of 25% of the list of member subscribers of American Psychological Association residing in the US (1243 questionnaires were sent out and 852 were usable). A second grouping consisted of foreign member subscribers and US member subscribers residing in foreign countries (157 sent out and 83 usable). A third grouping consisted of US non-APA member subscribers (63 sent out, 40 usable), a fourth of foreign non-APA member subscriber (75 sent out, 32 usable) and a fifth of student subscribers (199 sent out, 141 usable). The second part of the study "reviews the current operation and literature input of Psychological Abstracts, indicates areas of operation that require improvement, outlines a general design for the future operation of Psychological Abstracts, and describes the steps to be taken in achieving the design." Only descriptive data on Psychological Abstracts is presented in the second part.

Report#16: Innovations in scientific communication in psychology.	The study surveys two samples: "(1) Authors of the listed manuscripts and (2) Requestors who contacted the Authors for copies of the listed manuscripts." The first sample consists of 699 authors of which 603 returned usable questionnaires. The second sample consists of information provided by the authors on requests. The second part of the study polled samples of the presenters, requesters, attendants and immediate readers at the 73rd Annual Convention of the American Psychological Association 1965. These samples were compared with the answers of control groups (with the exception of the latter group). The number of questionnaires sent out ranged from 173 to 1164 (the total population only known for some of the groups) and the number of usable questionnaires ranged from 139 to 629. The response rate ranged from 54 to 90%.
Report#17: The use of scientific information in the undergraduate teaching of psychology.	The study surveyed "persons engaged in teaching psychology at undergraduate level". A sample from the faculties of 246 universities was sent a questionnaire. A total of 1123 questionnaires were sent out and 495 usable were returned.
Report#18: Information exchange at the 1966 American Psychological Association Annual Convention and the function of the convention. Proceedings in such exchange.	A sample of sessions from the 74th Annual Convention of the American Psychological Association 1966 was selected. A total of 10 invited addresses, 21 symposia and 23 contributed papers were included in the study. The study design was the same as in report#12 and the questions sought to analyse the success

	of proceedings as a communication medium.
Report#19: Information exchange activities involved in psychological work.	The study surveyed a sample of "APA members who were affiliated with specialized organizations [….] and a second sample who were not so affiliated". The first sample consisted of 20% of the population and 1904 questionnaires were sent out resulting in 1390 usable. The second sample consisted of about 10% of the unaffiliated APA members and 1726 questionnaires were sent out of which 1002 usable were returned.
Report#20: Scientific communication at the XVIII International Congress of Psychology, Moscow, 1966 and its implications for the design and operation of international meetings.	Data was collected through a series of simultaneous surveys from Amsterdam, Moscow and Washington and included information from both authors and attendants at the XVIII International Congress of Psychology, Moscow, 1966. Author surveys were sent out to app. 380 and 214 usable were returned. App. 490 questionnaires were sent out to attendants and 284 usable were returned.
Report#21: Networks of information communication among scientifically productive psychologists: An exploratory study.	To analyse specialised subgroups they selected "a group of persons, averaging about nine, in each of twelve areas of research generating publications in the 1959-63 period". Four areas were excluded on the basis of information from questionnaires and a total of 14 persons were interviewed. A second series of interviews or "rounds of communication" were conducted after the preparation of the initial draft of the report.

8. Intra-disciplinary differences in database coverage and the consequences for bibliometric research[4]

Tove Faber Frandsen[5] and Jeppe Nicolaisen

Royal School of Library and Information Science
Birketinget 6, DK-2300 Copenhagen S., DENMARK
tff@db.dk; jni@db.dk

Abstract

Bibliographic databases (including databases based on open access) are routinely used for bibliometric research. The value of a specific database depends to a large extent on the coverage of the discipline(s) under study. A number of studies have determined the coverage of databases in specific disciplines focusing on inter-disciplinary differences. However, little is known about the potential existence of intra-disciplinary differences in database coverage. Focusing on intra-disciplinary differences, the paper documents large database coverage differences within two disciplines (economics and psychology). The point extends to include both the uneven coverage of specialties and research traditions. The implications for bibliometric research are discussed, and precautions which need to be taken are outlined.

[4]Early results of this work were presented at the 3rd International COLLNET Conference, New Delhi, India, March 2007. This work was partly financed by a generous grant from the Nordic Research School in Library and Information Science (NORSLIS).
[5] Corresponding Author: Tove Faber Frandsen. Royal School of Library and Information Science. Birketinget 6, DK-2300 Copenhagen S. Phone: + 45 32 58 60 66, e-mail: tff@db.dk.

Introduction

The introduction of large bibliographic databases marks a significant development in the history of bibliometrics. Many branches of bibliometric research have grown out of, or been made possible by the use of these databases. However, the use of bibliographic databases for bibliometric research is not without its problems. According to Hood and Wilson (2003: 593), these problems may be seen as falling within three categories:

1. Errors or lack of consistency in the data (at the micro level)
2. Other types of problems and difficulties (at the macro level)
3. Problems with the tools that are made available by the database provider or host

One of the problems dealt with in the second category, is that of database coverage. This problem has both quantitative and qualitative aspects (Jacsó, 1997). The quantitative aspects concern among other things the size of the database(s), indexed document types, the number of English-language and foreign-language source documents, geographic coverage, and the time span and currency of the database(s). The qualitative aspects are partly about the inclusion of core journals and prestigious non-journal sources.

Using bibliographic databases for bibliometric research implies using these databases and their coverage as censuses of publications comparable to demographers using population censuses for demographic studies (White & McCain, 1989). Potential bias in the censuses will reflect itself in the results of studies based on these. Thus, the coverage of bibliographic databases has consequences for bibliometric research and this is valid regardless of specific choice of database (subject-specific databases, citation databases or databases based on open access resources).

It is crucial to recognize possible coverage problems before conducting bibliometric research. Using the citation indexes for research evaluation may produce quite biased and invalid results. Moed (2005) discusses the coverage problem and its implications for the use of the citation databases produced by the Institute of Scientific Information (ISI) in research evaluation. He presents a so-called *tentative* classification of disciplines according to their overall ISI coverage into three classes, with *excellent, good,* and *moderate coverage*, respectively. Disciplines in the *"excellent"* category

include molecular biology and biochemistry, biological sciences primarily related to humans, clinical medicine, physics and astronomy, and chemistry. The *"good"* category contains the disciplines of applied physics and chemistry, biological sciences primarily related to animals and plants, psychology & psychiatry. As well as other social sciences primarily related to medicine and health, geosciences, mathematics, engineering, and economics. The *"Moderate"* category contains other social sciences, and humanities and arts. Moed's classification system gives one the impression that whole disciplines are either excellent, good, or moderately covered in the citation databases, thus making research evaluation based on ISI data feasible to a varying degree. Although it is just a tentative classification system, it is nevertheless still too crude, as a division of disciplines into smaller units may reveal uneven coverage and thus casting the idea about *the* coverage of whole disciplines in to doubt. Unfortunately, there seems to be a lack of studies of the bibliographic coverage of specialties and research traditions within disciplines and its consequences for bibliometric research. This paper aims to close this gap. The focus is centered on intra-disciplinary differences. More specifically, the paper aims to investigate the bibliographic coverage of specialties and research traditions within disciplines and the consequences for bibliometric research. As such, the paper to some extent continues in the footsteps of previous studies. Yet, its narrow focus on intra-disciplinary differences distinguishes it from related studies focusing on, for instance, inter-disciplinary coverage, geographical coverage and coverage of publication types.

The paper is organized as follows: The next section provides an overview of related research. The following sections present the methods of data gathering and processing followed by results, discussions, and conclusions.

Related research

Disciplinary differences in publishing and citing behavior have been studied in various ways, but only seldom at the intra-disciplinary level. A search for studies focusing on database coverage issues at the intra-disciplinary level yielded no results. Yet, subject-specific database comparisons have been undertaken in a number of fields.

Inter-disciplinary and intra-disciplinary differences

Differences in publishing behavior can be analyzed on several levels. The analysis can be performed on a macro level as a comparison between e.g. the humanities, social sciences, natural sciences, and medicine as done by Kyvik (2003). The analysis can also be made at a more detailed level as done by Hyland (2000) examining the relationships between the cultures of eight disciplines and their unique discourses. Kling and McKim (2000) examine the heterogeneity of communications illustrated by an analysis of three disciplines. Knievel and Kellsey (2005) compare eight humanities fields. Kyvik (1988) focuses on six social sciences as he compares them with other fields and analyzes the differences among the social sciences. Lindholm-Romantschuk and Warner (1996) study the role of monographs in scholarly communication in philosophy, sociology and economics. Metcalfe (1995) illustrates the differences between disciplines by showing a difference in mean publication lag between two disciplines of 6.2 and 16.3 months. Nederhof et al. (1989) and Nederhof and Zwaan, (1991) analyze the composition of document types, their coverage by the citation indexes and the consequences for bibliometric research in six disciplines. An even more fine-grained analysis can be performed as shown by Bordons and Zulueta (1997) stressing that even within the same ISI heading differences are found between journals representing different research communities. Pharmacological teams and pharmacy teams are identified and their results show that the journals they submit their articles to for publication are very different. Hamilton (1990, 1991) shows that the un-citedness rate varies from 36.7 to 88.0% among fields and from 9.2 to 99.8 among sub-disciplines indicating that inter-disciplinary differences are not necessarily larger than intra-disciplinary differences. Laband (2002) compares conditions of co-authorships in economics and agricultural economics uncovering great differences in the author conditions.

Database coverage and bibliometric research

A number of researchers have investigated the coverage of a bibliographic database and the consequences for bibliometric research. The bibliometric consequences related to the problem of geographical coverage are well illustrated by Webster's (1998) analysis of a Polish sociological citation index (PSCI) and the Social Science Citation Index (SSCI). Her findings strongly imply that bibliometric indicators based on SSCI paint

one picture of Polish sociology and the PSCI another. Another study by Narvaez-Berthelemot and Russell (2001) contains an analysis of 4,326 social science journals from the UNESCO DARE-database. It reveals that 64% of production of journals in the world takes place in high income countries. Furthermore, that SSCI primarily consists of journals from the rich countries (97%). Apart from that there is a smaller group of journals from middle income countries and finally there is a very small group of journals from low income countries (0.7%). All countries apart from the US have fewer journals included in SSCI than the UNESCO DARE-database. Bordons, Fernandez and Gomez (2002) report on some of the problems for peripheral countries in relation to calculations of journal impact factors, and stress that it should be borne in mind that large parts of the scientific output in these countries are not included in the citation indexes. Studies on of database coverage are also made using open access data sources. Hajjem, Harnad and Gingras (2005) analyze coverage of 1,307,038 articles from 10 disciplines from 1992-2003 by open access based resources and find an overall percentage of OA articles ranging from 5 to 16% depending on discipline, year and country. Swan et al. (2005) have studied the coverage of different document types by open access resources and find some document types to be better covered by open access resources. The bibliometric consequences related to the problem of document type coverage are evident in the study by Cronin, Snyder and Atkins (1997). The three authors constructed a database comprising 30,000 references from 90 books randomly chosen among those reviewed in top sociological journals. They compare lists of the 26 authors most cited in the books and in the top 24 sociology journals, and their findings demonstrate that there are two distinct populations of highly cited authors in sociology: One consisting of authors cited in the journal literature, another of authors cited in the monographic literature. Given the citation databases' limited coverage of monographic citing material, the latter population may regularly go unrecognized. Finally, the number of databases needed to cover the literature on a specific topic has been found to vary considerably. Hood and Wilson (2001) report that for a typical search topic, the single most inclusive database covers 23 to 37% of the relevant literature. To cover 95% of the literature in their study of 14 topics would require the use of 11 to 35 databases. Hood and Wilson (2001) study topics from various fields. Other researchers have examined database coverage of particular subject areas. Walters and Wilder (2003) provide a comprehensive overview of this research.

Methods

The focus of this study is on the coverage of bibliographic databases and the consequences for bibliometric research with a narrow focus on the disciplines of economics and psychology. Both are classified as well-covered in Moed's classification system (Moed, 2005: 138), however, a more detailed analysis is needed. This may be accomplished by dividing the disciplines into research traditions or specialties. All disciplines embrace a number of (often competing) research traditions that to some extent are distributed among the specialties that shape the whole discipline. What characterizes a specialty is, according to Meadows (1998), the phenomenon or phenomena, which members of the specialty study. Laudan (1977) invokes the idea of a large-scale unit in science that he calls a *research tradition*. A research tradition is held together by common ontological assumptions about the nature of the world and methodological principles about how to revise theories and develop new ones. Research traditions are consequently not the same as specialties. A research tradition is "a set of ontological and methodological do's and don'ts" (Laudan, 1977: 80) whereas a specialty is a specific part, fraction or division of a larger discipline.

The specialties within the discipline of economics were determined using EconLit. EconLit is the American Economic Association's electronic bibliography of economics literature. EconLit contains abstracts, indexing, and links to full-text articles in economics journals. It abstracts books and indexes articles in books, working papers series, and dissertations. EconLit indexes the economics literature using EconLit Subject Descriptors, which is comparable to headings in the JEL Classification System (www.econlit.org). The JEL Classification System is a classification developed for the economics literature by the Journal of Economic Literature and widely used in the discipline. Barrett, Olia and Von Bailey (2000) also use the JEL Classification System to show that economics is a discipline characterized by great specialization. Other databases could have been used as the benchmark database, but EconLit was chosen because of the subject descriptors.

The year 1991 was chosen as the starting point for the analyses because the required information was not available for the previous years. A 15 year publication period (1991-2005) was employed. On the basis of the JEL classification system the following

four specialties were selected: Health economics, mathematical and quantitative methods, economic history and schools of economic thought and methodology. The varying publication patterns were analyzed using the JEL Classification System in EconLit. All publications indexed with the classification code of the specialty were ranked according to publication type year by year. An overview of the document composition is available in appendix 1 to 4. The same publications were also ranked according to journal name, resulting in 15 annual lists for each specialty of journals publishing one or more articles indexed in EconLit with the classification code of the specialty. Subsequently, the 60 lists of journals resulting in a total of 34,496 journal articles were scrutinized and checked for indexing in the citation databases (Social Sciences Citation Index, Science Citation Index, Arts & Humanities Citation Index) for each of the examined years. The investigation was made on journal level and not article level, implying that each article was not looked up in the indexes, but the journal was. This implies that if only a selection of the articles in a journal is indexed in the citation indexes, it is possible that the specific article is not indexed but as the citation indexes normally include all research articles of a journal (Moed, 2005: 113) the possible bias of this procedure is assessed to be low.

To give a preliminary answer to the problem concerning the consequences of coverage, a study of the relative sizes of the four specialties was conducted. The relative size of specialties was analyzed using different sources for performing research evaluation. As this analysis includes open access sources an analysis of the 15 years is not possible because the content of open access based databases is not static and thus the most recent year in the analysis (2005) was selected. The point of reference was EconLit, the citation databases and a delineation in the citation databases to the top 20 journals within the economics subject category. The top 20 journals were measured by journal impact factor (JIF) as available through the 2005 JCR social sciences edition in the subject category economics (isiknowledge.com/jcr). An overview of the 20 journals is available in appendix 5. The rationale for the analysis performed on the 20 journals with the highest JIF is a number of previous studies that have used a similar delineation as the sampling method (e.g. Hodgson & Rothman, 1999; Kocher & Sutter, 2001; Frost et al., 2003) or as means to characterize the quality of publications (Klaić & Klaić, 2004). Furthermore, a tool for citation analysis based on open access resources was included. The publications were located using Google Scholar - an alternative to the existing citation databases (Noruzi, 2005; Bakkalbasi et al., 2006; Kousha & Thelwall, 2006;

Neuhaus & Daniel, 2007). Only journal articles were included in order to make the analysis comparable to the one in the citation databases. Unlike the citation indexes, open access based resources do not allow systematic analyses of neither the indexing policy nor the consequences of it. The indexing policies of services based on open access resources are difficult to analyze. This is partly caused by a lack of available information of the indexing policy (e.g. Google Scholar) and partly because the service providing access to the data is not in control of the indexing policy (e.g. archives based on authors self-archiving their work). Consequently, the influence on bibliometric studies of the indexing policy is even more difficult to investigate as we are limited by the options available through the services and as Neuhaus and Daniel (2007) state:

> "Google Scholar currently processes its sources in an unsystematic, unpredictable and fragmentary manner. For lack of adequate options for browsing, searching and saving results in structured output formats it is difficult to make even elementary bibliometric analyses efficiently."

Consequently, all 4,230 journal articles had to be looked up individually in Google Scholar. Noruzi (2005) outlines the search techniques available through Google Scholar. In this case the queries submitted were based on words from the title and author's last name. If the query did not yield any results, the number of words were initially increased and afterwards decreased. It had to be a total match to be registered as available in full text via Google Scholar. This implies that a journal article had to be available as pre-print or post-print. This could lead to a decrease in the shares of publications retrieved as OA. But there is a risk of vast differences between e.g. a working paper and the subsequent journal article so the distinction had to be made. Furthermore, it was checked if the publication was available in full text or if only the bibliographic information was available. Some links are to toll access journals and others to open access databases that may not include access to full text. RePEc (Research Papers in Economics) is an example of a decentralized database of working papers, journal articles and software components freely available. However, as stated on the website (http://repec.org), RePEc does not contain full-text journal articles as RePEc services provide links to many full text articles. Yet, a personal or institutional subscription to follow those links is often required.

In order to be able to study differences at the level of research traditions, three research traditions were chosen from the psychological specialty psychotherapy and psychotherapeutic counselling:

- cognitive therapy,
- behavior therapy
- psychoanalytic therapy.

These three are different research traditions because they hold different ontological assumptions about psychological phenomena as well as different ideas about how to study them (Robins, Gosling & Craik, 1999; Nicolaisen, 2004: chapter 5). To assess the variations in coverage of various databases caused by intra-disciplinary differences, the specialties within the selected discipline of psychology must be determined. For that purpose we used the database PsycINFO that indexes the literature in psychology and related behavioral and social sciences, including psychiatry, sociology, anthropology, education, pharmacology, and linguistics. Records from 1967 and beyond are indexed using the controlled vocabulary from the Thesaurus of Psychological Index Terms.[6] The publications of three research traditions were determined using the Thesaurus of Psychological Index Terms. An overview of the composition of document types and the total publications is available in appendix 6 to 8. Note that due to the indexing policy of PsycINFO the shares of working papers are not available in these appendices.

The same 15 year publication period (1991-2005) was employed, and the varying publication patterns were analyzed using the Thesaurus of Psychological Index Terms in PsycINFO. This resulted in 45 lists of journal names comprising 16,193 publications. All publications indexed with the classification code of the research tradition were ranked according to publication type year by year. The same publications were also ranked according to journal name. Again, the list of journals was scrutinized and checked for indexing in the citation databases. The study of the relative sizes of the three research traditions was performed using the same method as the study of economics specialties. The only difference was the list of top 20 journals measured by JIF which was determined by merging the 10 subject categories related to psychology.

[6] The three index terms used are *cognitive therapy, behavior therapy & behavior modification*, and *psychoanalytic therapy*.

An overview of the 20 journals is available in appendix 9. In Google Scholar all 1,366 articles were looked up individually.

Results

In the following we show the development in the size of a selection of document types. Each figure depicts the moving averages[7] of a specific document type of the total research output within a specialty from 1991 to 2005 in intervals of three years (although the first and last year are only in intervals of two years).

Figure 1 shows the significance of journal articles within the four selected specialties. It should be noted that this document type includes all types of journal articles (e.g. reviews, research articles and notes).

[7] A moving average is simply the average of a series of numbers over a period of time which is constantly updated by dropping the oldest value and then adding the newest value and recalculating the average. Using moving averages smooth a data series and make it easier to spot trends.

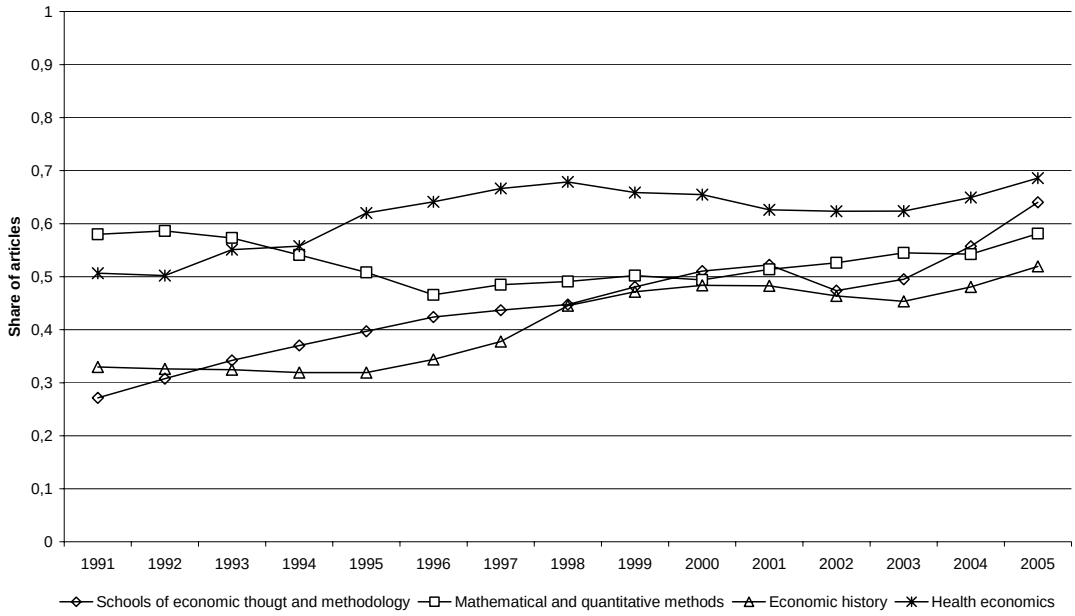


Figure 1. Share of journal articles of total output in four economics specialties in EconLit 1991-2005

As is quite evident, the journal article is of growing importance in all the specialties, and it increases from shares of 25 to 56% in 1991 to shares of 57 to 71% in 2005. The remaining document types are primarily books and working papers. The results showing the importance of these document types are available in appendix 1-4. The relative size of the book seems to be relatively stable in some disciplines on a level of about 2 to 7% of the research output. However, within one specialty it appears as if books are losing their importance. In the specialty of economic history the book is rapidly decreasing in relative size over the years, although, the book is still at a much higher level in this specialty than in the other three. The working paper is a document type with an increasing significance within all four specialties. However, the importance of the working papers is varying considerably among the specialties as mathematical and quantitative methods hold a share of 30% in 2005 whereas the other specialties have shares varying from 4 to 12% in 2005.

Based on this analysis we can conclude that specialties within the discipline of economics have quite varying publication patterns, and we will now examine the implications of these findings for the coverage in the citation databases. Figure 2 illustrates the coverage of journal articles in the citation indexes.

Figure 2 depicts the shares of journal articles covered by the citation indexes. It should be noted that the influence of publication patterns is excluded as only journal articles are included. The coverage varies from 40 to more than 90% of the journal articles.

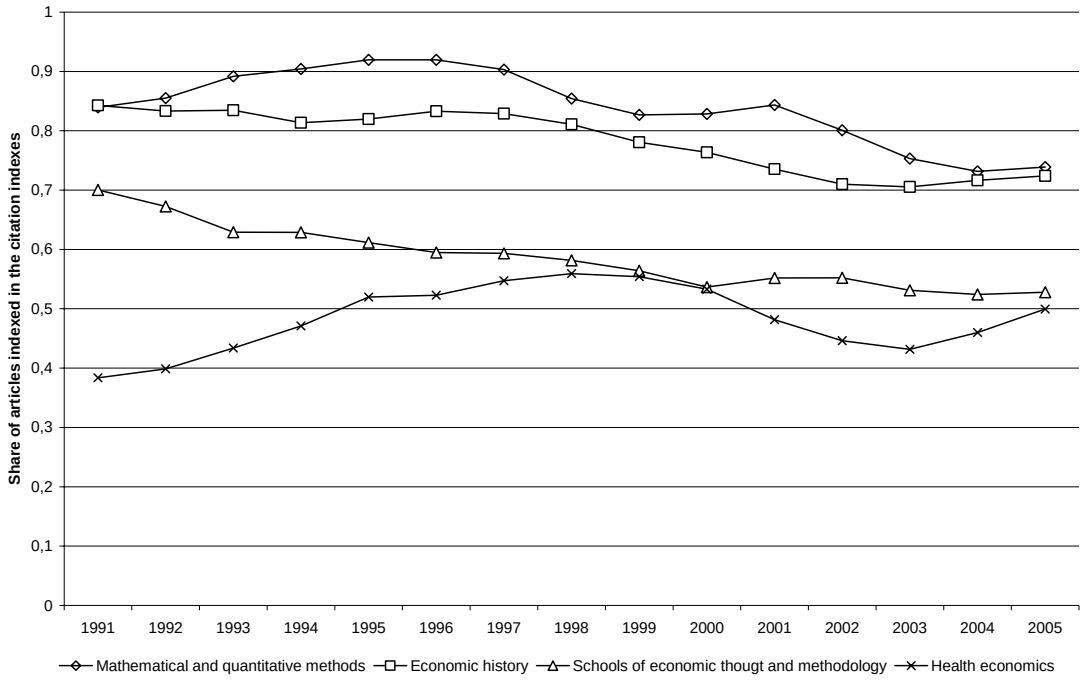


Figure 2. Share of journal articles indexed in the citation databases from 1991 to 2005 of four economics specialties.

Obviously, in a research evaluation these varying degrees of coverage can hypothetically imply that some specialties appear less productive than others. However, the central issue is to what extent the varying degrees of coverage influence the results of citation analyses and research evaluation. To give a preliminary answer, we have conducted a study of the relative sizes of the four specialties in 2005. The results are presented in table 1.

	EconLit	ISI citation databases	Top 20 journals	Google Scholar
Mathematical and quantitative methods	36	38	21	50
Economic history	17	17	13	19
Schools of economic thought and methodology	19	15	07	13
Health economics	28	30	60	18

Table 1. Relative sizes of four economics specialties in per cent: journal articles published in 2005.

Table 1 shows the relative sizes of the four specialties vary considerably when using different pools of documents. It should be noted that there is no "true" relative size among these four pools of documents as they are all determined by their indexing policy. In EconLit mathematical and quantitative methods make up 36% of the total amount of journal articles produced by these four specialties. Economic history is the smallest amounting to 17%. Schools of economic thought & methodology and health economics are represented by respectively 19 and 28%. In an evaluation performed using EconLit of productivity measured by the number of journal articles this would be their relative sizes. The same analysis done by using the citation databases would depict a somewhat different picture. Economic history would hold the same relative size whereas health economics and mathematical and quantitative methods would have slightly bigger shares. However, this increase in size is associated with a decrease for schools of economic thought and methodology which would appear to be a less productive area than e.g. economic history although in EconLit it was the other way around. Turning to the 20 journals with the highest JIF, the relative sizes are considerably different from the two previous pools of documents. The share of health economics doubles (and becomes the largest specialty) and the rest of the specialties lose shares (although they do not lose shares equally). Finally, if the analysis had been performed using Google Scholar, health economics turns into one of the three smallest specialties whereas mathematical and quantitative methods becomes the dominating specialty by far. Summing up the table, it is evident that these four pools of documents are not duplicating the same picture of productivity in these four economics specialties.

Taking the analysis one step further, we analyze three research traditions within one specialty as we look at three research traditions of psychology: Cognitive therapy, behavior therapy & behavior modification and psychoanalytic therapy.

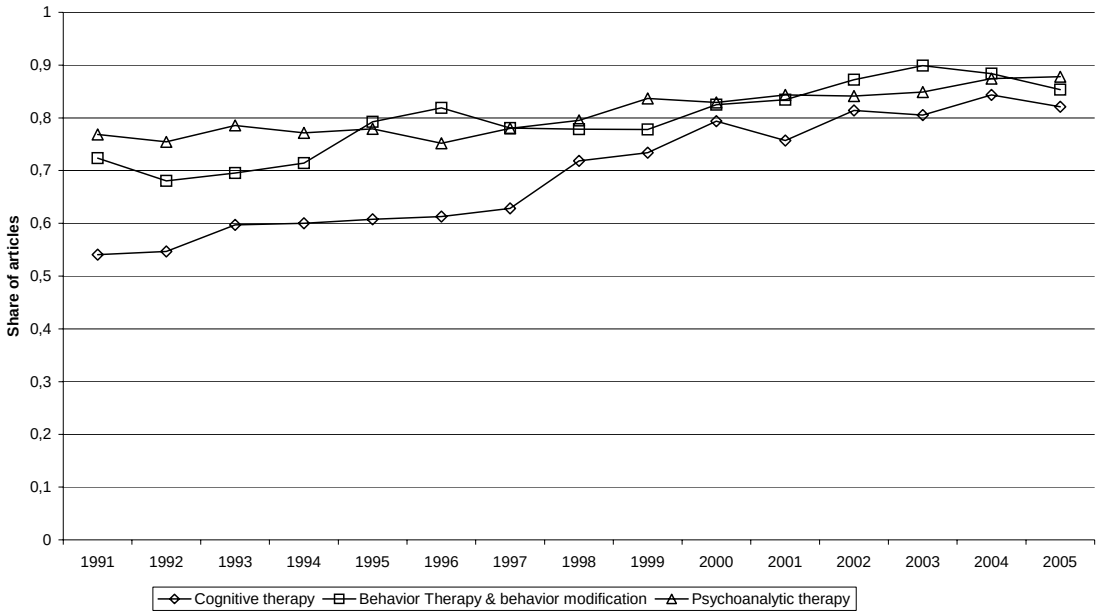


Figure 3. Share of journal articles in three Psychological research traditions in PsycINFO 1991-2005.

Figure 3 shows the significance of journal articles within the three selected research traditions. Again, it should be noted that this document type includes all types of journal articles (e.g. reviews, research articles and notes). Compared to the four economics specialties the journal article is much more important and is also of growing importance in all the research traditions as it increases from shares of 51% to 74% in 1991 to shares of 81% to 91% in 2005.

The three research traditions have relatively similar publication patterns, during the last 5 or 6 years of the period and the publication patterns cannot be used to explain differences in visibility. Their visibility in a research evaluation is to a large extent

dependent upon the indexing policy of the tools used for the research evaluation. As can be seen in figure 4, the coverage of journal articles in ISI varies considerably across research traditions and these three research traditions are thus not indexed equally well each year.

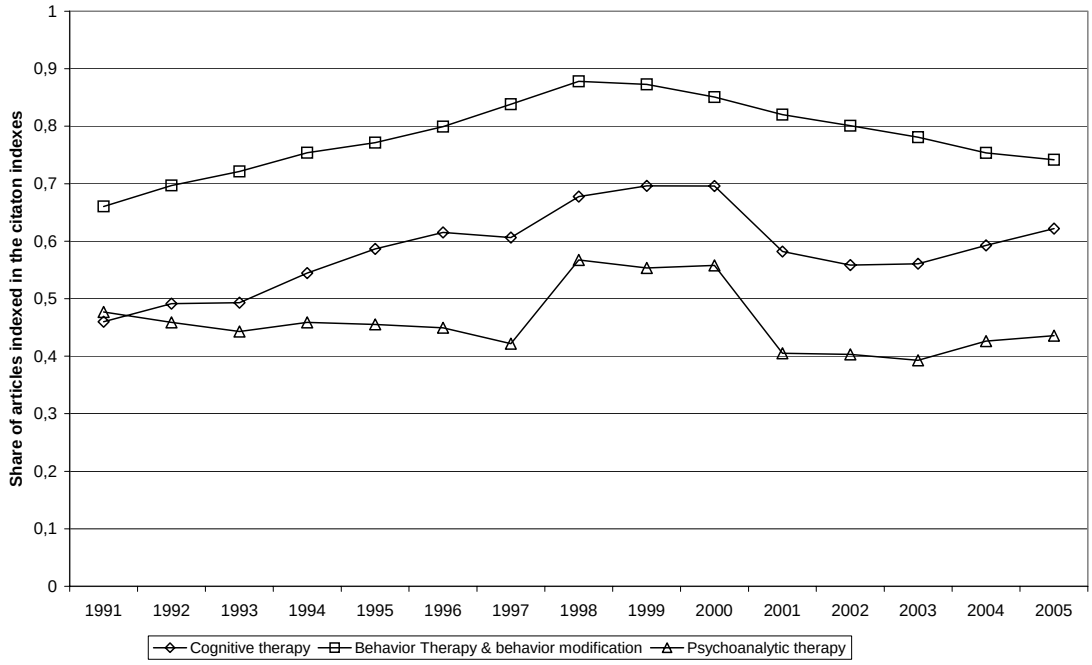


Figure 4. Share of journal articles indexed in the citation databases from 1991 to 2005 of three Psychological research traditions.

Throughout the entire period, behavior therapy & behavior modification is considerably better covered by the citation indexes compared to the other two research traditions in general, and psychoanalytical therapy in particular. The poor coverage of the latter is somewhat surprising as this research tradition has its own subject category in the citation indexes (*Psychoanalysis*).

Turning to the implications of the uneven coverage of research traditions, table 2 provides an overview of the relative sizes of the three research traditions.

	PsycINFO	ISI citation databases	Top 20 journals	Google Scholar
Cognitive therapy	31	36	69	41
Behavior therapy & behaviour modification	14	20	28	22
Psychoanalytical therapy	54	44	03	37

Table 2. Relative sizes of three Psychological research traditions in per cent: journal articles published in 2005.

Like the specialties within economics, the relative sizes of the three research traditions also vary considerably using different pools of documents. Again it must be stressed that there is no "true" relative size among these four pools of documents. In PsycINFO psychoanalytical therapy provides a little over 50% of the journal articles of these three research traditions. Cognitive therapy produces 31% of the journal articles and behavior therapy & behaviour modification the remaining 14%. The same analysis, using the citation databases as pool of documents, produces a different picture. Psychoanalytical therapy loses 10 percentage points whereas the other two research traditions share the "profit" almost equally. However, the ranking of research traditions by size does not change. The same cannot be said about the relative sizes using the 20 highest ranking JIF journals. Psychoanalytical therapy almost vanishes and cognitive therapy becomes the clearly identifiable leading research tradition. Finally, using Google Scholar involves lost shares for psychoanalytical therapy reducing it to the second largest research tradition. Using Google Scholar does, however, depict a picture of two strong research traditions and a third somewhat smaller research tradition. The problems of uneven coverage of specialties within economics are therefore also evident when it comes to research traditions in psychology. The four pools of documents are not replicating the same picture of productivity in the three Psychological research traditions.

Discussion

The results of our empirical study show that a number of specialties in the discipline of economics and research traditions in the discipline of psychology are not represented equally well in the databases. As the findings of the present study only relate directly to the disciplines of economics and psychology, we cannot assume they aply to other disciplines. However, though restricted to these disciplines the results do have broader implications.

Using a different method, Moed (2005) evaluates the coverage of the citation indexes in order to lay the groundwork for a proper understanding of the performance measures computed on the basis of ISI data. His study includes an analysis of the coverage of economics in which he finds the percentage of references to documents published in journals, relative to total references in 2002 to be 56 (Moed, 2005: 129). Furthermore, he finds the ISI coverage of journals within economics to be 83% leading to an overall ISI coverage of 47. However, these figures can be further differentiated as we find the importance of the journal as a publishing medium to range from 52 to 63% of the total output. Furthermore, we find the ISI coverage of journal articles to range from 58 to 83%. This leads to an overall ISI coverage of 30, 36, 46 and 48%. This variation of values indicates that some specialties within economics are what Moed refers to as well covered although others are only moderately covered.

In psychology and psychiatry Moed (2005: 130) finds the percentage of references to documents published in journals, relative to total references in 2002 to be ranging from 69 to 81 (psychology and psychiatry is divided into sub-disciplines). Furthermore, he finds the ISI coverage of journals to be 86 and 91% leading to an overall coverage in psychology and psychiatry of 60 to 73%. However, these figures can be further differentiated when looking at research traditions within psychology as we find the importance of the journal as publishing medium to range from 74 to 86% of the total output. Furthermore, we find the ISI coverage of journal articles to range from 39 to 85%. This leads to overall coverage of 33, 41 and 73%. In the three examined research traditions the variation is even greater than in the four economic specialties. Some psychological research traditions are well covered whereas others are just moderately covered.

According to Moed (2005: 140), the degree of coverage of a field determines the type of research assessment study necessary to perform an adequate analysis. The moderately covered fields should not be analysed relying on ISI data alone, but require supplementary analyses based on data not available in the ISI databases. In some of the moderately covered fields it may not even be possible to perform citation analyses. The research assessment study needs to be adjusted according to the field as studies based solely on ISI data at risk of being biased in moderately covered fields.

It is easy to imagine how bibliometric studies based on an uneven coverage of a discipline's specialties and research traditions, could produce biased or invalid results. Normally a distinction is made between two kinds of bibliometric studies. The first concerns studies based on publication analysis. The second concerns studies based on citation analysis.

Publication analyses normally seek to measure and compare the scientific output of authors, institutions and countries. This is usually accomplished by counting the number of publications indexed in databases. It is thus of utmost importance that databases used for publication analyses cover all specialties and research traditions of the analyzed disciplines adequately. Otherwise the bias in the coverage will immediately reflect itself in the results of the publication analysis, thus invalidating its conclusions.

Bias will also reflect itself in the results of citation analyses. There are four main applications of citation analysis (Zunde, 1971; Nicolaisen, 2007):

1. Qualitative and quantitative evaluation of scientists, publications and scientific institutions
2. Modeling of the historical development of science and technology
3. Information search and retrieval
4. Knowledge organization based on bibliographic coupling and co-citation analysis

Authors tend to cite authors from the same specialty and/or research tradition (Nicolaisen, 2004). Uneven database coverage of specialties and research traditions will consequently affect the results of all four applications. The volume of citations to the

well covered specialties and research traditions will be disproportionate higher than the volume of citations to the ill covered specialties and research traditions.

Another problem with uneven database coverage of a discipline's specialties and research traditions concerns the issue of sampling. The majority of bibliometric studies are based on retrieved data from databases. The databases are normally used for two related purposes: 1. for selecting a sample for further analysis, and 2. for detecting the publication output of the sample units and/or how many times the sample units are cited. Blind reliance on uneven database coverage, when selecting a sample for further analysis, is clearly problematic. Such a sample may at best be regarded a fractionized sample, and any results based on such a sample has limited generalizability (Nicolaisen, 2006).

When conducting bibliometric studies it is crucial to identify possible coverage problems that may lead to biased results. To recognize such problems the analyst must be knowledgeable about the discipline(s) under study. It is vital to be aware of various specialties and research traditions within the discipline(s), and to examine their coverage in the databases selected for studies. It may be possible to compensate for uneven database coverage, but only if the analyst knows *what* to normalize for.

Conclusion

Intra-disciplinary differences in database coverage affect the results of bibliometric research based on retrieved data from databases. We have documented significant differences in the disciplines of economics and psychology, and revealed quite uneven coverage of economic specialties and psychological research traditions. These observable facts have consequences for all bibliometricians - not only those studying the disciplines of economics and psychology. Intra-disciplinary differences in database coverage could very well be found in other disciplines as well. Consequently, specialties and research traditions of any discipline cannot be assumed to be covered equally well in the databases. It is important to be aware of this and to take appropriate precautions before initiating bibliometric studies using bibliographic databases.

Acknowledgements: The authors wish to thank the anonymous referees for their constructive suggestions for improvements and Peter Joseph Woods for proof-reading the manuscript.

References

Bakkalbasi, N.; Bauer, K.; Glover, J. & Wang, L. (2006). Three options for citation tracking: Google Scholar, Scopus and Web of Science. http://eprints.rclis.org/archive/00006080/02/GS_Scopus_WoS_04182006_preprint.pdf. [Downloaded 05-30-2007].

Barrett, C.B; Olia, A. & Von Bailey, D. (2000). Subdiscipline-specific journal rankings: Whither Applied Economics. *Applied Economics*, 32: 239-252.

Bordons, M.; Fernandez, M.T. & Gomez, I. (2002). Advantages and limitation in the use of impact factor measures for the assessment of research performance in a peripheral country. *Scientometrics*, 53(2): 195-206.

Bordons, M. & Zulueta, M. A. (1997). Comparison of research team activity in two biomedical fields. *Scientometrics*, 40(3), 423-436.

Cronin, B., Snyder, H. & Atkins, H. (1997). Comparative citation rankings of authors in monographic and journal literature: A study of sociology. *Journal of Documentation,* 53(3): 263-273.

Frost, S.; Murphy, R.; Webster, P. & Schmidt, U. (2003). Are top journals biased against eating disorders topics? *American Journal of Psychiatry*, 160(2), 363–365.

Gallagher, L.A., ed. (2007). Thesaurus of psychological index terms. 11th ed. Washington, DC: *American Psychological Association.*

Hajjem, C., Harnad, S. & Gingras, Y. (2005). Ten-Year Cross-Disciplinary Comparison of the Growth of Open Access and How it Increases Research Citation Impact. IEEE Data Engineering Bulletin 28(4), 39-47.

Hamilton, D. P. (1990). Publishing by - and for? - the numbers. *Science,* 250, 1331-32

Hamilton, D. P. (1991). Research papers: who's uncited now. *Science*, 251, 25.

Hodgson, G. & Rothman, H. (1999). The editors and authors of economics journals: a case of institutional oligopoly? The Economics Journal, 109(2), 165-186.

Hood, W.W. & Wilson, C.S. (2001). The scatter of documents over databases in different subject domains: How many databases are needed? *Journal of the American Society for Information Science*, 52(14): 1242-1254.

Hood, W.W. & Wilson, C.S. (2003). Informetric studies using databases: Opportunities and challenges. *Scientometrics*, 58(3): 587-608.

Hyland, K. (2000). Disciplinary discourses: Social interactions in academic writing. Harlow, UK: Pearson Education Ltd.

Jacsó, P. (1997). Content evaluation of databases. *Annual Review of Information Science and Technology*, 32: 231-267.

Klaic, Z. B. & Klaic, B. (2004). Croatian scientific publications in top journals according to the Science Citation Index for the 1980-2000 period. *Scientometrics*, 61(2), 221-251.

Kling, R. & McKim, G. (2000). Not Just a Matter of Time: Field Differences in the Shaping of Electronic Media in Supporting Scientific Communication. *Journal of the American Society for Information Science*, 51(14), 1306-1320.

Knievel, J.E., & Kelley, C. (2005). Citation analysis for collection development: A comparative study of eight humanities fields. *The Library Quarterly* 75(2): 142-168.

Kocher, M & Sutter, M. (2001). The Institutional concentration of authors in top journals of Economics during the last two decades. *Economic Journal*, 111(5), 405-421.

Kousha, K. & Thelwall, M. (2007). Google Scholar Citations and Google Web/URL Citations: A Multi-Discipline Exploratory Analysis. *Journal of the American Society for Information Science and Technology*, 58(7), 1055-1065.

Kyvik, S. (1988). Internationality of the social sciences: the Norwegian case. *International Social Science Journal*, 115, 163-172.

Kyvik, S. (2003). Changing trends in publishing behaviour among university faculty, 1980-2000. *Scientometrics*, 58(1), 35-48.

Laband, D. N. (2002). Contribution, attribution and the allocation of intellectual property rights: economics versus agricultural economics. *Labour economics,* 9, 125-131.

Laudan, L. (1977). *Progress and its Problems: Toward a Theory of Scientific Growth*. Berkeley, CA: University of California Press.

Lindholm-Romantschuk, Y. & Warner, J. (1996), The role of monographs in scholarly communication: an empirical study of philosophy, sociology and economics. *Journal of Documentation*, 52(4), 389–404.

Metcalfe, N. B. (1995). Journal impact factors. *Nature*, 77, 260-261

Moed, H.F. (2005). *Citation Analysis in Research Evaluation*. Dordrecht, NL: Springer.

Narvaez-Berthelemot, N. & Russell, J.M. (2001). World Distribution of Social Science journals: A view from the periphery. *Scientometrics,* 51(1): 223-239.

Nederhof, A. J., Zwaan, R.A., De Bruin, R.E., & Dekker, P. J. (1989). Assessing the usefulness of bibliometric indicators for the humanities and the social sciences: A comparative study. *Scientometrics,* 15, 423-435.

Nederhof, A. J. & Zwaan, R. A. (1991) Quality judgments of journals as indicators of research performance in the humanities and the social and behavioral sciences. *Journal of the American Society for Information Science and Technology*, 42(5, 332-340.

Neuhaus, C. & Daniel, H-D. (2007). Data sources for performing citation analysis: An overview. *Journal of Documentation*, 64(2), 193-210. [Downloaded 06-01-2007].

Nicolaisen, J. (2004*). Social Behavior and Scientific Practice – Missing Pieces of the Citation Puzzle* [PhD-thesis]. Copenhagen, DK: Royal School of Library and Information Science.

Nicolaisen, J. (2006). Traditional author co-citation analysis: A discussion of the sampling problem. *Proceedings of the 1st International Conference on Multidisciplinary Information Sciences and Technologies, InSciT2006:* 635-639.

Nicolaisen, J. (2007). Citation analysis. *Annual Review of Information Science and Technology*, 41: 609-641.

Noruzi, A. (2005). Google Scholar: The new generation of citation indexes. *LIBRI*, 55(4): 170-180. http://eprints.rclis.org/archive/00005595/01/Google_Scholar%2C_The_New_Generation_of_Citation_Indexes.pdf. [Downloaded 06-01-2007].

Robins, R.W., Gosling, S.D. & Craik, K.H. (1999). An empirical analysis of trends in Psychology. *American Psychologist*, 54(2): 117-128.

Swan, A., Needham, P., Probets, S., Muir, A., Oppenheim, C., O'Brien, A., Hardy, R. and Rowland, F. (2005). Delivery, Management and Access Model for E-prints and Open Access Journals within Further and Higher Education. Technical Report, JISC, HEFCE. http://eprints.ecs.soton.ac.uk/11001/01/E-prints_delivery_model.pdf. [Downloaded 06-01-2007].

Walter, W. H. & Wilder, W. (2003). Bibliographic index coverage of a multidisciplinary field. *Journal of the American Society for Information Science and Technology*, 54(14): 1305–1312.

Webster, B.M. (1998). Polish sociology citation index as an example of usage of national citation indexes in scientometric analysis of social sciences. *Journal of Information Science,* 24(1): 19-32.

White, H.D. & McCain, K.W. (1989). Bibliometrics. *Annual Review of Information Science and Technology*, 24: 119-186.

Zunde, P. (1971). Structural models of complex information sources. *Information Storage and Retrieval*, 7: 1-18.

Appendix 1. Publications of the specialty: Schools of economic thought and methodology.

Year	Total publications	Journal articles		Books		Working papers		Other publication types	
1991	1687	427	(25)	124	(7)	6	(0)	1130	(67)
1992	1644	476	(29)	116	(7)	2	(0)	1050	(64)
1993	1243	473	(38)	75	(6)	7	(1)	688	(55)
1994	1462	521	(36)	110	(8)	2	(0)	829	(57)
1995	1263	472	(37)	87	(7)	9	(1)	695	(55)
1996	1245	574	(46)	80	(6)	12	(1)	579	(47)
1997	1303	570	(44)	75	(6)	12	(1)	646	(50)
1998	1525	628	(41)	101	(7)	14	(1)	782	(51)
1999	1336	659	(49)	80	(6)	13	(1)	584	(44)
2000	1290	692	(54)	69	(5)	26	(2)	503	(39)
2001	1447	726	(50)	87	(6)	18	(1)	616	(43)
2002	1112	588	(53)	78	(7)	20	(2)	426	(38)
2003	1228	480	(39)	68	(6)	24	(2)	656	(53)
2004	1238	700	(57)	67	(5)	27	(2)	444	(36)
2005	1131	809	(72)	69	(6)	50	(4)	203	(18)

Percentages are shown in brackets.

Appendix 2. Publications of the specialty: Mathematical and quantitative methods

Year	Total publications	Journal articles		Books		Working papers		Other publication types	
1991	1609	902	(56)	71	(4)	67	(4)	569	(35)
1992	1685	1009	(60)	60	(4)	115	(7)	501	(30)
1993	1643	984	(60)	58	(4)	158	(10)	443	(27)
1994	1714	894	(52)	70	(4)	70	(4)	680	(40)
1995	1765	886	(50)	58	(3)	217	(12)	604	(34)
1996	2428	1217	(50)	65	(3)	396	(16)	750	(31)
1997	3080	1214	(39)	102	(3)	408	(13)	1356	(44)
1998	2478	1386	(56)	73	(3)	480	(19)	539	(22)
1999	2755	1431	(52)	74	(3)	488	(18)	762	(28)
2000	3114	1332	(43)	88	(3)	882	(28)	812	(26)
2001	2603	1392	(53)	81	(3)	462	(18)	668	(26)
2002	2016	1167	(58)	69	(3)	408	(20)	372	(18)
2003	1742	809	(46)	62	(4)	528	(30)	343	(20)
2004	2460	1456	(59)	52	(2)	597	(24)	355	(14)
2005	2664	1521	(57)	77	(3)	824	(31)	242	(9)

Percentages are shown in brackets.

Appendix 3. Publications of the specialty: Economic history

Year	Total publications	Journal articles		Books		Working papers		Other publication types	
1991	1331	413	(31)	161	(12)	14	(1)	743	(56)
1992	1103	385	(35)	148	(13)	19	(2)	551	(50)
1993	1232	393	(32)	173	(14)	21	(2)	645	(52)
1994	1304	399	(31)	204	(16)	29	(2)	672	(52)
1995	1319	438	(33)	198	(15)	31	(2)	652	(89)
1996	1479	472	(32)	182	(12)	48	(3)	777	(53)
1997	1410	536	(38)	194	(14)	49	(3)	631	(45)
1998	1399	607	(43)	189	(14)	49	(4)	554	(40)
1999	1186	619	(52)	140	(12)	39	(3)	388	(33)
2000	1474	677	(46)	162	(11)	94	(6)	541	(37)
2001	1294	608	(47)	151	(12)	72	(6)	463	(36)
2002	1130	586	(52)	133	(12)	78	(7)	333	(29)
2003	1314	529	(40)	152	(12)	80	(6)	553	(42)
2004	1656	726	(44)	132	(8)	110	(7)	688	(42)
2005	1187	713	(60)	119	(10)	143	(12)	212	(18)

Percentages are shown in brackets.

Appendix 4. Publications of the specialty: Health economics

Year	Total publications	Journal articles	Books	Working papers	Other publication types
1991	625	413 (44)	161 (4)	14 (1)	743 (56)
1992	694	385 (58)	148 (5)	19 (1)	551 (50)
1993	709	393 (49)	173 (6)	21 (3)	645 (52)
1994	900	399 (58)	204 (4)	29 (2)	672 (52)
1995	889	438 (60)	198 (4)	31 (4)	652 (49)
1996	995	472 (68)	182 (3)	48 (3)	777 (53)
1997	1121	536 (65)	194 (3)	49 (4)	631 (45)
1998	1224	607 (67)	189 (2)	49 (4)	554 (40)
1999	1346	619 (72)	140 (3)	39 (5)	388 (33)
2000	1564	677 (59)	162 (3)	94 (7)	541 (37)
2001	1573	608 (66)	151 (2)	72 (5)	463 (36)
2002	1255	586 (63)	133 (3)	78 (7)	333 (29)
2003	1281	529 (58)	152 (5)	80 (8)	553 (42)
2004	1718	726 (66)	132 (3)	110 (8)	688 (42)
2005	1670	713 (71)	119 (3)	143 (11)	212 (18)

Percentages are shown in brackets.

Appendix 5. The 20 highest JIF ranking economics journals in JCR 2005

Journal name	JIF
Quarterly Journal of Economics	4.775
Journal of Economic Literature	4.054
Journal of Economic Geography	3.222
Journal of Health Economics	2.708
Journal of Economic Perspectives	2.634
Econometrica	2.626
Journal of Economic Growth	2.577
Journal of Financial Economics	2.385
Journal of Political Economy	2.245
Brookings Papers on Economic Activity	2.118
Journal of Risk and Uncertainty	2.100
Review of Economic Studies	2.035
Health Economics	1.919
Journal of Accounting & Economcis	1.877
American Economic Review	1.806
Economic Geography	1.757
Journal of International Economics	1.667
Journal of Monetary Economics	1.661
Journal of Law & Economics	1.609
Feminist Economics	1.595

Appendix 6. Publications of the research tradition: Cognitive therapy

Year	Total publications	Journal articles		Books		Other publication types	
1991	179	91	(51)	64	(36)	24	(13)
1992	193	112	(58)	58	(30)	23	(12)
1993	221	124	(56)	69	(31)	28	(13)
1994	223	147	(66)	58	(26)	18	(8)
1995	225	131	(58)	83	(37)	11	(5)
1996	238	138	(58)	90	(38)	10	(4)
1997	223	149	(67)	69	(31)	5	(2)
1998	298	188	(63)	107	(36)	3	(1)
1999	103	88	(85)	14	(14)	1	(1)
2000	302	217	(72)	79	(26)	6	(2)
2001	359	291	(81)	68	(19)	0	(0)
2002	382	283	(74)	96	(25)	3	(1)
2003	446	397	(89)	45	(10)	4	(1)
2004	553	431	(78)	111	(20)	11	(2)
2005	493	424	(86)	59	(12)	10	(2)

Percentages are shown in brackets.

Appendix 7. Publications of the research tradition: Behavior Therapy & behavior modification

Year	Total publications	Journal articles		Books		Other publication types	
1991	348	247	(71)	52	(15)	49	(14)
1992	263	195	(74)	21	(8)	47	(18)
1993	381	225	(59)	91	(24)	65	(17)
1994	268	201	(75)	40	(15)	27	(10)
1995	246	197	(80)	42	(17)	7	(3)
1996	279	232	(83)	42	(15)	5	(2)
1997	273	227	(83)	44	(16)	2	(1)
1998	254	173	(68)	79	(31)	2	(1)
1999	95	78	(82)	16	(17)	1	(1)
2000	245	203	(83)	32	(13)	10	(4)
2001	269	221	(82)	48	(18)	0	(0)
2002	297	252	(85)	45	(15)	0	(0)
2003	289	275	(95)	12	(4)	2	(1)
2004	227	204	(90)	20	(9)	3	(1)
2005	239	194	(81)	41	(17)	4	(2)

Percentages are shown in brackets.

Appendix 8. Publications of the research tradition: Psychoanalytic therapy

Year	Total publications	Journal articles	Books	Other publication types
1991	790	585 (74)	182 (23)	23 (3)
1992	793	634 (80)	143 (18)	16 (2)
1993	959	700 (73)	240 (25)	19 (2)
1994	930	772 (83)	140 (15)	18 (2)
1995	878	667 (76)	202 (23)	9 (1)
1996	909	682 (75)	209 (23)	18 (2)
1997	786	590 (75)	189 (24)	7 (1)
1998	766	643 (84)	115 (15)	8 (1)
1999	357	282 (79)	71 (20)	4 (1)
2000	778	677 (87)	93 (12)	8 (1)
2001	775	635 (82)	139 (18)	1 (0)
2002	816	685 (84)	131 (16)	0 (0)
2003	922	802 (87)	120 (13)	0 (0)
2004	945	794 (84)	142 (15)	9 (1)
2005	812	739 (91)	57 (7)	16 (2)

Percentages are shown in brackets.

Appendix 9. The 20 highest JIF ranking psychology journals in JCR 2005

Journal name	JIF
Behavioral and Brain Sciences	9.885
Annual Review of Psychology	9.784
Psychological Bulletin	9.746
Trends in Cognitive Sciences	9.155
Psychological Review	7.986
Advances in Experimental Social Psychology	7.000
American Psychologist	6.460
Monographs of the Society for Research in Child Development	5.667
Journal of Experimental Psychology – General	5.242
Journal of Clinical Psychiatry	5.038
Psychotherapy and Psychosomatics	4.966
Journal of Cognitive Neuroscience	4.533
Psychological Science	4.502
Journal of Abnormal Psychology	4.383
Journal of Personality and Social Psychology	4.211
Neuropsychologia	4.119
Journal of the American Academy of Child and Adolescent Psychiatry	4.113
Neurobiology of Learning and Memory	4.091
Journal of Consulting and Clinical Psychology	4.023
Cognitive Psychology	3.932

9. The integration of open access journals in the scholarly communication system: Three science fields

Tove Faber Frandsen [8]
Royal School of Library and Information Science
Copenhagen S. Denmark
tff@db.dk

Abstract:

The greatest number of open access journals (OAJs) is found in the sciences and their influence is growing. However, there are only a few studies on the acceptance and thereby integration of these OAJs in the scholarly communication system. Even fewer studies provide insight into the differences across disciplines. This study is an analysis of the citing behaviour in journals within three science fields: biology, mathematics, and pharmacy and pharmacology. It is a statistical analysis of OAJs as well as non-OAJs including both the citing and cited side of the journal to journal citations. The multivariate linear regression reveals many similarities in citing behaviour across fields and media. But it also points to great differences in the integration of OAJs. The integration of OAJs in the scholarly communication system varies considerably across fields. The implications for bibliometric research are discussed.

[8] The author is greatly indebted to Jeppe Nicolaisen and Birger Hjørland. Furthermore, the author acknowledges the valuable comments by an anonymous referee.

Introduction

Studies show that the influence of OAJs in the scholarly communication system is growing. Kling and Callahan (2003) provide an overview of studies on perception of OA journals. The study by McVeigh (2004) documents that the number of OAJs in the citation indexes provided by ISI Thomson™ is growing, both in terms of creating new titles and conversion of established titles. Furthermore, OAJs are dominantly lower-ranking journals in their field measured by Journal Impact Factor (JIF) and Immediacy Index although OAJs rank higher by Immediacy Index, than by JIF. Sotudeh and Horri (2007a) analyse the performance of OAJs in terms of expected citation rates and conclude that OA is widely recognised by scientific communities.

However, the influence of OAJs is not growing with the same rate in all fields as field differences are indicated in the existing literature in terms of the number and the acceptance of OAJs. The sciences are undeniably leaders in establishing OAJs, however, they are distributed unevenly within the sciences (Borgman, 2007: 186). In addition, Kling and McKim (2000) conclude that:

> "[C]ommunicative plurality and communicative heterogeneity are durable features of the scholarly landscape, and […] we are likely to see field differences in the use of and meaning ascribed to communications forums persist, even as overall use of electronic communications technologies both in science and in society as a whole increases." (Kling & McKim, 2000: 1306)

The study by McVeigh (2004) shows great field differences, and high-ranking OA journals are the most likely to be found in the fields of physics, engineering and mathematics.

The present study aims to contribute to the understanding of the implications of the increasing number of scientific publications published by journals running under an OA model. The influence of OAJs is typically measured by their number or share; however, there are no bibliometric studies on whether they are integrated or recognised generally in the scholarly communities. Another aspect yet to be investigated is whether the

integration of OAJs in both OAJs and non-OAJs varies from field to field. More specifically, this paper addresses the following research question: are OAJs integrated in the journal communication system? The research question can be specified further: Does the citing of OA and non-OA journals depend on the citing journal being OA or non-OA and do field differences influence that citing behaviour?

The paper is structured as follows: The following section formulates an operational definition of OAJs to be used in this study. The next section presents the collected data and the chosen methods, followed by a presentation of the results of the analysis. Furthermore, the paper includes a discussion, and the last section contains conclusions and the perspectives of the paper.

Open access journals

Open access means that scientific publications are made freely available on the Internet, without any access restrictions. OA can be achieved using a number of different financing models. Thus, Willinsky (2006) identifies ten different models of providing open access to scholarly publications including both self-archiving and OAJs. OAJs can be seen as the second phase or strategy in the process of achieving open access (Brody & Harnad, 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).

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). This study focuses on OAJs; however, non-OAJs and OAJs can in practise be difficult to separate as non-OAJs can de facto be at least partially OA. Journals managed under other business models than

OA can be partly OA because single publications can be self-archived by e.g. the author(s). Some journals provide free access after an embargo period. Furthermore, many scholars have full text access to non-OAJ articles through university or corporate licences implying that these scholars would not perceive non-OAJ and OAJ differently in terms of access.

The focus of this study is on a potential keenness or reluctance of authors (or editors) to integrate OAJs in the reference lists of the accepted publications in non-OAJs and OAJs. The key issue is not the accessibility of single publications but on the perception of quality tied to OAJs in various disciplines. Thus, self-archived non-OAJ articles do not distort the point made here. However, the vast field differences in use of OA and choice of OA model do necessitate that the fields are analysed separately.

The operational definition of OAJs in this study is journals managed under a business model that does not charge readers or their institutions for access. Journals with limited free access (e.g. free access is restricted to a select period of time or a select sample of publications in the journal) are not regarded as OAJs.[9]

Methods

A statistical analysis of the importance of type of media (OA versus non-OA) on the citing and cited side is an approach similar to the one used by Baldi (1997, 1998) on document level.[10] Using multiple linear regressions on both cited and citing journals enables controlling for different characteristics of the journals as well as for their degree of interaction or dependency. The method has been used by Frandsen (2005), albeit with a focus on the degree of dependency between American and European economics journals.

[9] Ulrich's Periodicals Directory™ bases their distinction of OAJs and non-OAJs on a similar definition.

[10] Although it is remarkably complex to determine what citations measure (the reader is referred to Nicolaisen, 2007 for recent review of theories of citation analysis), the citation analysis framework provides the opportunity to analyse what Cronin (2001: 2) refers to as "links (reference citations) provided routinely by authors in their reports and papers [which] are a means of exposing the underlying socio-cognitive structure of science."

The journals included in this study were selected from three science disciplines. Ideally, social sciences and humanities disciplines could have been included but as relatively high numbers of OAJs within the selected disciplines are required, the present study is restricted to the sciences. The data in this study was analysed as three separate datasets as it is crucial to be able to control for discipline and sub-discipline specific variances. The analysis was performed on the basis of data from 2006 as this was the most recent publication year completed in the citation databases at the time of the data collection. The disciplines selected were biology, mathematics, and pharmacy and pharmacology as described by the classification scheme of Ulrich's Periodicals Directory™. They were selected on the basis of the number of OAJs which varies considerably among disciplines.[11] For the statistical analyses it is essential to select a discipline with a relatively high number of OAJs. A considerable number of OAJs convert into non-OAJs (Sotudeh & Horri, 2007b), however, as the development over time is not the focus in this study it is sufficient that the journal had OA status at the time of data collection. An overview of the included journals and some of their characteristics is provided in appendix 1-3. Within the discipline of biology journals from 5 biology sub-disciplines were selected resulting in 74 journals, within mathematics 25 general journals, and within pharmacy and pharmacology 20 journals from 2 sub-disciplines were selected. The smaller number of journals in the two latter disciplines is due to a smaller number of OAJs in these two disciplines and resulted in the exclusion of language as independent variable in those models.

The variables in the linear regressions consisted of a dependent variable and a number of independent variables. The dependent variable in this study was the dependency of the citing journal on the cited journal. Dependency was measured by the number of references from one journal to another which was determined using the Science Citation Index (SCI). The cited work field is uncontrolled and consequently attention must be paid to the different forms of names as well as articles in press. However, as the total number of references and the number of references to other publication types than journal articles (e.g. books and working papers) varies considerably across journals[12],

[11] The number of OAJs is presently about 2700 according to the Directory of Open Access Journals (DOAJ, http://www.doaj.org). However, this number can be divided into disciplines showing significant differences in the number of OAJs.

[12] Some journals have more references to other document types such as monographs and working papers and an increased dependency on these document types should be reflected in

normalisation is needed. The number of references is normalised by dividing the number of references by the total number of references in the citing journal and multiplying by a hundred to get the relative dependency in per cent. The relative dependency of journal *i* on journal *j* is defined as follows:

$$\text{Relative dependency}_{i,j} = \frac{\text{Number of references}_{i,j} * 100}{\text{Number of references}_{i}}$$

The distribution of references across journals is expected to be influenced by a number of factors not related to the issue of OA. These factors were sought captured by a number of independent variables as they could potentially distort the results if not included. The relative dependency of journals was primarily described by the following independent variables: sub-discipline, JIF, publication patterns, OA and variables describing the relationship of the citing and cited journal. The sub-disciplinary variables consisted of dichotomous variables of the sub-disciplines (a journal could belong to more than one sub-discipline within the discipline as it depended on the indexing of the journals in Ulrich's). Geographical relations were described by a variable containing the geographic location of authors i.e. the share of authors located in three regions: North America, Western Europe and the rest of the world.[13] Furthermore, variables concerning the languages of the journals were constructed. A variable concerning the languages of the journals were constructed. Information on the language of an article was available in the citation indexes and a variable was created describing the language of the publication as being English or non-English. Information on the JIF and total number of citations received by the journals was available in Journal Citation Reports (JCR). It is essential to control for the average number of citations received per publication in any analysis of citing behaviour on journal level. Some journals receive 10 or 100 times as many citations as other journals and much higher degrees of dependencies on such journals must be expected. This is important as the focus of this study is not on explaining why some journals receive more citations than others, but rather to focus on the importance of the OA status of the cited journals in the citing journal. A variable

lower dependencies on the journals in this data sample. The importance of the journal article is varying considerably across disciplines (Moed, 2005: 129-130) and sub-disciplines (Frandsen & Nicolaisen, 2008)

[13] In this study the general division of regions by the World Trade Organisation (WTO) is applied.

describes the share of reviews (as journals consisting of many reviews are expected to be less likely to cite other journals also consisting of many reviews). The dichotomous OA variable was constructed on the basis of information from Ulrich's and confirmed on the journals' websites. Finally, a number of variables described the dyadic character of the relationship between journals. The variables sought to capture the effect of own group preference which is an effect detected on many levels. Self-citations is a strong own group preference as confirmed by e.g. Fassoulaki et al. (2000), Aksness (2003), Frandsen (2005) and Frandsen (2007). Other, but probably weaker, own group variables are variables describing similarities between the citing and cited journal (e.g. same sub-discipline). The variable, dependency on this data sample, is the combined relative dependencies of a journal to all the journals in the dataset. The degree a journal depends on the other journals in this data sample should, in principle, increase the dependency on each single journal in the sample.

The results of the analysis presented below consist of different statistical analyses of the data material. One of the variables mentioned above was not included in the final models as it did not contribute to the understanding of the dependent variable (language). The slope coefficients for the linear relationships are given. Pearson's r^2 reveals information about the degree of correlation between the dependent and the independent variables when controlling for the effects of the other variables.

Results

In the following focus will be upon the variable characterizing whether the cited journal is an OAJ or not. If the coefficient to this variable in the statistical analysis is positive it indicates that this field has shown citing behaviour that gives more citations to OAJs than would have been expected on the basis of their characteristics. Thus such a field we shall describe as "*OA including*". If the coefficient to the variable is negative it oppositely indicates that journals in this field are citing OAJs less than their characteristics would imply, and such a field we shall denote as "*OA excluding*". Finally, if the coefficient is insignificantly different from zero, we shall denote the field as "*OA neutral*" as no apparent difference in the dependence of OA and non-OA can be found.

Table 1 shows the results of a multivariate linear regression analysis of the citing behaviour of all journals from the samples within the three disciplines. It is apparent that the citing behaviour of all three fields is relatively well described by such an analysis, as R squared exceeds one half in all three regressions.

Table 1. Multivariate linear regression analysis of citing behaviour of all journals. The dependent variable is relative dependency in per cent.

Variable	**Biology**	**Mathematics**	**Pharmacy and pharmacology**
Intercept	-0.0140	-0.3175	
Cited journal OA			-0.1078
Dependency on this data sample	0.0085	0.0492	0.0506
Cited journal JIF	0.0094	0.1825	0.0336
Share of authors from Western countries	-0.0307		
Indicator for journal self-citations	1.2037	1.7419	1.4534
Difference in JIF between citing and cited journal	-0.0013		
Belong to same sub-discipline	0.0528	-	0.1381
Combined share of reviews	-0.0407		-0.1884
R squared	0.504	0.536	0.542
# of observations	5476	625	400

Numbers not shown are not significant at the .05 significance level. The sign "-" indicates that this variable was not included in the analysis for Mathematics.

From table 1 is it clear that when looking at the disciplines as a whole we should denote biology and mathematics as OA neutral whereas pharmacy and pharmacology is OA excluding. Turning briefly to the other variables in the statistical models we see that there is a rather large degree of homogeneity in the importance of these across the three fields. In all three cases it is thus the case that the variables “Dependency on this data sample”, “Cited journal JIF” and “Indicator for journal self-citations” influence the citing behaviour in the same (expected) way. It is thus to be expected that a higher JIF will tend making the degree of dependency higher. Similarly, the degree a journal depends on the other journals in this data sample should increase the dependency on

each single journal in the sample – at least on average. Finally, it is well-known that a large percentage of citations are journal self-citations. In this context this translates into a higher degree of dependency on a journal when this is actually the journal itself. Since the dependency variable is measured in per cent, the coefficients to the indicator for self-citations show that the share of self-citing is between 1.2 and 1.7 percentage points higher than to a journal with otherwise similar characteristics. The variable "Combined share of reviews" is significant in two of the analyses with a negative sign. The reason for is that journals publishing a large share of review to a smaller extent are dependent on each other than on other types of journal where the reviewed literature is published. The importance of sub-discipline is also remarkable, although already established in the existing literature by e.g. Bordons and Zulueta (1997); Frandsen and Nicolaisen (2008).

Coming back to our main variable of interest, namely the OA indicator variable, the significant coefficient to this variable in the regression for pharmacy and pharmacology means that an average OAJ received 8 citations less from each of the other journals than a similar non-OAJ in the year 2006. As already pointed out, this type of analysis cannot provide explanations of low or high levels of dependencies of a single journal across the data sample implying that the OAJs of OA including and excluding fields probably receive relatively more citations from journals not included in the study or outside their field. The focus of the study is on the analysis of OAJs and non-OAJs separately.

In order to analyze whether the citing behaviour is different for OAJs and non-OAJs, an extra two sets of multivariate regressions have been carried out, where the OAJs and non-OAJs have been analyzed separately. The results of these analyses are shown in tables 2 and 3.

Table 2. Multivariate linear regression analysis of citing behaviour of non-OAJs. The dependent variable is relative dependency in per cent.

Variable	**Biology**	**Mathematics**	**Pharmacy and pharmacology**
Intercept	-0.0187	-0.3453	
Cited journal OA		0.1193	-0.1165
Dependency on this datasample	0.0100	0.0496	0.0468
Cited journal JIF	0.0077	0.2002	0.0316
Share of authors from Western countries	-0.0329		
Indicator for journal self-citations	1.3578	2.1000	1.2963
Difference in JIF between citing and cited journal			
Belong to same sub-discipline	0.0498	-	0.1218
Combined share of reviews	-0.0423		-0.1630
R squared	0.560	0.596	0.598
# of observations	3404	400	240

Numbers not shown are not significant at the .05 significance level. The sign "-" indicates that this variable was not included in the analysis for Mathematics.

From table 2 it is apparent that when we look only at traditional non-OAJs we still find that pharmacy and pharmacology is OA excluding and biology is OA neutral. However, contrary to the overall status of mathematics, non-OAJs in this field are OA including, giving more citations to OAJs in the sample than to other similar journals. In terms of numbers this means that the non-OAJs in mathematics gave 2 citations extra to each of the OAJs in the sample. The coefficients to the control variables are almost identical to those in table 1, confirming the overall robustness of the method.

Table 3. Multivariate linear regression analysis of citing behaviour including OAJs. The dependent variable is relative dependency in per cent.

Variable	**Biology**	**Mathematics**	**Pharmacy and pharmacology**
Intercept		-0.3159	
Cited journal OA	0.0248		
Dependency on this datasample	0.0050	0.0485	0.0481
Cited journal JIF	0.0231	0.1545	0.0840
Share of authors from Western countries			
Indicator for journal self-citations	0.9237	1.1151	1.6710
Difference in JIF between citing and cited journal	-0.0148		
Belong to same sub-discipline	0.0612	-	0.1684
Combined share of reviews	-0.0517		-0.2948
R squared	0.428	0.473	0.517
# of observations	2072	225	160

Numbers not shown are not significant at the .05 significance level. The sign "-" indicates that this variable was not included in the analysis for Mathematics.

From table 3 we can see that for OAJs the analysis gives rise to somewhat different results than the two previous tables. We thus see that within biology OAJs are OA including whereas OAJs within mathematics, and pharmacy and pharmacology apparently are neutral with respect to their citing behaviour towards other OAJs. The coefficient to the OA variable for biology corresponds to one extra citation to each OAJ from each OAJ compared to the number of citations to a similar non-OA journal.

Table 4 summarizes the findings with respect to citation behaviour towards OAJs for the three analyzed disciplines. As was already apparent in the presentation of the statistical analyses above there are great field differences in the integration of OAJs

Table 4. Summary of OA inclusion in three sciences and subdivisions hereof.

	Overall	**Non-OA**	**OA**
Biology	0	0	+
Mathematics	(+)	+	0
Pharmacy and pharmacology	-	-	0

+ indicates OA inclusion, - indicates exclusion while 0 indicates neutrality. Sign shown in parentheses indicates significance at the .1 significance level.

The main conclusion to be drawn from the table thus seems to be that any statement indicating a uniform advantage or disadvantage for OAJs is questionable. In fact the analysis in this paper seems to imply that the status of OAJs and the subsequent citations to them in OAJs as well as non-OAJs depend greatly on the fields and subfields in question.

Discussion

Before addressing the consequences of the findings it must be emphasized that the analysis in this paper only has included a subset of the science disciplines. The results of the empirical study are based on references from three science disciplines and cannot necessarily be generalized to other fields. Furthermore, self-archiving makes the demarcation of OAJs and non-OAJs vague. Finally, it should be noted that this type of analysis cannot provide explanations of low or high levels of dependencies of a single journal across the whole data sample as this effect (to a large extent) is captured by the variable containing JIF values. The method can, however, explain lower or higher dependencies of a set of journals within a field. However, although limited, the results have implications for bibliometric studies.

To a wide extent OAJs and non-OAJs can be described by the same elements. Many of the same variables in the multiple linear regressions are statistically significant with identical signs providing evidence of the strength of the models used in this study. Both OAJs and non-OAJs cite journals with a high JIF more and journals depending greatly on this data sample have higher dependencies themselves as cited journals. Regardless

of being OA or NOA the journals have strong own group dependencies in terms of self-citations and sub-discipline self-citations. These variables are included as control variables and expected to turn out statistically significant with a positive coefficient. However, the results of the analyses of OAJs and non-OAJs differ in terms of the use of OAJs depending on the discipline.

The present study contributes to the understanding of the so-called open access postulate defined as "authors are more likely to read, and thus cite, articles that are made available under an OA model" (Craig et al., 2007). The findings in this study indicate that if such an effect exists for OAJs it is probably not found in all disciplines. The development of OA is not just a matter of the number of OAJs in a field but also to what extent they are accepted and used in non-OAJs as well as OAJs.

This study gives insight into the developments in scholarly communication. 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. As Van Raan (1997: 447) states:

> "In our opinion, the new electronic publishing developments will not influence conceptually [the] main functions of scientific communication. Of course, technology will certainly influence, even dramatically, these functions, particularly in terms of performance and of mechanical improvement."

Furthermore, 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. Following Barjak (2006) there are two major concerns:

- Causation
- Distinguishing between function and technology.

It must be stressed that one cannot make any causal arguments on the basis of the present analysis as it can point to the underlying structure of OAJs and non-OAJs in various fields but not explain why there are varying patterns of interaction among journals. It could be an issue of author perceptions of OA. Findings by Swan & Brown (2005) show that the main reason for not having published in an OAJ is lacking familiarity of the concept of OAJs or with specific OAJs in their field. Authors who have not published in OAJs perceive them to have low prestige and impact, directly in contrast to the perception of authors who have published in an OAJ. Another possible explanation is the specific research areas within sub-disciplines as Zhao (2005) and Talja, Savolainen and Maula (2004) find publishing behaviour being closely related to the research area of the author. A third perspective is the issue of access. Authors publishing in OAJs could have less access to articles published in non-OAJs and they must thus depend more on the publications freely available on the Internet in e.g. OAJs. However, it could also be self-archived publications which are beyond the scope of this analysis to investigate.

On the other hand the second concern is easily determined as this study is based on data material with the same function. Both OAJs and non-OAJs are publishing mediums and the difference in citing behaviour is thus to be found in the financing models or the culture surrounding the journals. Summing up, it should be emphasized that until the causality has been further investigated one should be careful making definite conclusions on the integration of OAJs in the scholarly communication system. However, it can be concluded that there are differences in the scholarly communication in OAJs and non-OAJs across fields.

The results have implications for all researchers conducting bibliometric studies and the consequences for bibliometric research could be widespread. It will affect individuals or groups under study, how the data source used for the evaluation is covering OAJs and non-OAJs. Bibliometric studies can be made using a wide variety of data sources and

perhaps a combination of several data sources is preferable (Zhao & Strotmann, 2007). Regardless, the choice of data source(s) the relative share of OAJs can make a difference for the evaluation of individuals or institutions as it is related to a different citing behaviour within some fields. Should a pool of documents be sampled for further analysis, it is obviously of great importance how this is done. Such a sample could be problematic in terms of the relative size of OAJs and non-OAJs represented in the sample and one should be careful making generalizations (Nicolaisen, 2006).

To ensure valid results of bibliometric studies it is crucial to recognize possible biases in coverage in terms of the access aspect that may lead to biased results. In an evaluation that takes place across a wide board of journals (being both OA and non-OA) attention must be paid to the factors that may be determining the results of the analysis and appropriate precautions must be taken before initiating bibliometric studies using journal articles from either one or both as pools of documents. Analyses into the underlying structures of a discipline provide valuable insight in the scholarly communication of that field. Journal interaction analysis as performed in this study can reveal some of the hidden structures that are determinants for the results from citation analysis.

Conclusion

The results of this study confirm the many similarities between OAJs and non-OAJs which are in accordance with the resemblance in function. However, the results also point to dissimilarities. Causation cannot be determined in the present study; however, it is clear that in some fields authors publishing in OAJs are demonstrating different citing behaviour than authors publishing in non-OAJs. Within biology the non-OAJs are OA neutral with respect to their citing behaviour towards other OAJs. The OAJs within biology are OA including, giving more citations to OAJs in the sample than to other similar journals. Within pharmacy and pharmacology the non-OAJs as well as the OAJs are OA excluding, giving less citations to OAJs in the sample than to other similar journals. Finally, within mathematics non-OAJs are OA including whereas OAJs are neutral. Even within OAJs there is no guarantee of acceptance and integration of OAJs in general on the level we would expect based on a comparison with non-OAJs with similar characteristics.

Literature

Aksnes, D.W. (2003). A macro study of self-citations. Scientometrics, 56(2), 235-246.

Baldi, S. (1997). A network approach to the analysis of citation flows: A comparative study of two research areas in the natural and the social sciences. Department of Sociology, Ohio State University, Colombus.

Baldi, S. (1998). Normative versus social constructivist processes in the allocation of citations: a network analytic model. American Sociological Review, 63, 829-846.

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.

Bordons, M. & Zulueta, M. A. (1997). Comparison of research team activity in two biomedical fields. Scientometrics, 40(3), 423-436.

Borgman, C. L. (2007). Scholarship in the digital age: information, infrastructure, and the Internet. Cambridge, MA: MIT Press.

Brody, T. & Harnad, S. (2005). Keynote Lecture: Providing Open Access to peer-reviewed articles to maximise and measure their research impact. Online: http://www.oai.unizh.ch/symposium/docs/Harnad.ppt.

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.

Cronin, B. (2001). Bibliometrics and beyond: some thoughts on web-based citation analysis. Journal of Information Science, 27(1), 1-7.

Fassoulaki, A.; Paraskeva, A.; Papilas, K. & Karabinis, G. (2000). Self-citations in six anaesthesia journals and their significance in determining the impact factor. British Journal of Anaesthesia, 84(2), 266-269.

Finholt, T. (2002). Collaboratories. Annual Review of Information Science and Technology, 36, 73-107.

Frandsen, T. F. (2005). Journal interaction: a bibliometric analysis of economics journals. Journal of Documentation, 61(3), 385-401.

Frandsen, T. F. (2007). Journal self-citations - analysing the JIF mechanism. Journal of Informetrics, 1(1), 47-58.

Frandsen, T. F., Nicolaisen, J. (2008). Intra-disciplinary differences in database coverage and the consequences for bibliometric research. Journal of the American Society for Information Science and Technology.

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.

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.

Kling, R. & Callahan, E. (2003). Electronic Journals, the Internet, and Scholarly Communication. Annual Review of Information Science and Technology (ARIST), 37,127-77.

Kling, R. & McKim, G. (2000). Not Just a Matter of Time: Field Differences in the Shaping of Electronic Media in Supporting Scientific Communication. Journal of the American Society for Information Science, 51(14), 1306-1320.

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.

McVeigh, M.E. (2004). Open Access Journals and the ISI Citation Database: Analysis of Impact
Factors and Citation Patterns, Thomson Scientific Whitepaper. www.thomsonisi.com/media/presentrep/essayspdf/openaccesscitations2.pdf.

Moed, H.F. (2005). Citation Analysis in Research Evaluation. Dordrecht, NL: Springer.

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.

Nicolaisen, J. (2006). Traditional author co-citation analysis: A discussion of the sampling problem. Proceedings of the International Conference on Multidisciplinary Information Sciences and Technologies, InSciT2006: 635-639.

Nicolaisen, J. (2007). Citation analysis. Annual Review of Information Science and Technology, 41: 609-641.

Mackenzie Owen, J. (2007). The Scientific Article in the Age of Digitization. Dordrecht: Springer.

Sotudeh, H. & Horri, A. (2007a). The citation performance of open access journals: A disciplinary investigation of citation distribution models. Journal of the American Society for Information Science and Technology, 58(13), 2145-2156.

Sotudeh, H. & Horri, A. (2007b). Tracking open access journals evolution: Some considerations in open access data collection validation. Journal of the American Society for Information Science and Technology, 58(11), 1578-1585.

Swan, A. & Brown, S. N. (2004) JISC/OSI Journal Authors Survey Report. http://www.jisc.ac.uk/uploaded_documents/JISCOAreport1.pdf. [Downloaded 07-30-2007].

Talja, S.; Savolainen, R. & Maula, H. (2005). Field differences in the use and perceived usefulness of scholarly mailing lists. Information Research, 10(1), paper 200.

Van Raan, A.F.J. (1997). The future of the quality assurance system: its impact on the social and professional recognition in the era of electronic publishing. Journal of Information Science 23(6), 445-450.

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. Cambridge, Mass.: MIT Press. XV, 287p.

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.

Zhao, D. & Strotmann, A. (2007). Can citation analysis of web publications better detect research fronts? Journal of the American Society for Information Science and Technology, 58(9), 1285-1302.

Appendix 1: Biology journals included in the study

	OA	Dependency on this data sample	Share of reviews	Share of authors from North America	Share of authors from Western countries
Acta Biochimica Polonica	1	4,0	0,1	0,1	0,2
Acta Bioquimica Clinica Latinoamericana	1	1,9	0,1	0,0	0,1
Acta Protozoologica	1	3,2	0,0	0,0	0,5
Acta Zoologica	1	0,8	0,1	0,0	0,3
Advances in Biochemical Engineering - Biotechnology	0	1,9	0,0	0,9	1,0
Advances in Carbohydrate Chemistry and Biochemistry	0	1,2	0,7	0,1	0,7
African Zoology	0	1,3	0,0	0,1	0,4
American Journal of Hematology	0	2,6	0,0	0,4	0,6
American Journal of Primatology	0	2,7	0,0	0,7	0,8
American Museum Novitates	1	4,0	0,0	0,8	0,8
Animal Genetics	0	2,8	0,0	0,2	0,7
Annual Review of Genetics	0	2,1	1,0	0,5	1,0
Applied Biochemistry and Microbiology	0	4,2	0,0	0,0	0,0
Archives of Biochemistry and Biophysics	0	4,4	0,0	0,5	0,8
Archives of Microbiology	0	6,2	0,0	0,2	0,6
Behavior Genetics	0	2,3	0,0	0,5	0,9
Biochemistry	0	3,8	0,0	0,6	0,8
Biological Chemistry	0	4,0	0,1	0,3	0,8
Bioprocess and Biosystems Engineering	0	5,1	0,0	0,2	0,5
Bioscience, Biotechnology, and Biochemistry	1	5,5	0,0	0,0	0,0
BMC Biotechnology	1	6,2	0,0	0,2	0,8
BMC Evolutionary Biology	1	3,6	0,0	0,3	0,8
BMC Genetics	1	3,4	0,0	0,4	0,8
BMC Genomics	1	5,0	0,0	0,3	0,8
BMC Microbiology	1	5,5	0,0	0,3	0,7
Brazilian Archives of Biology and Technology	1	10,1	0,0	0,0	0,1
Brazilian Journal of Microbiology	1	4,8	0,0	0,0	0,1

Bulletin of the American Museum of Natural History	1	1,2	0,9	0,4	0,6
Canadian Journal of Microbiology	0	4,9	0,0	0,4	0,6
Caribbean Journal of Science	1	1,5	0,0	0,6	0,6
Clinical Biochemistry	0	2,0	0,1	0,4	0,6
Clinical Microbiology and Infection	0	3,5	0,1	0,1	0,8
Contributions to Zoology	1	2,2	0,1	0,1	0,8
Current Microbiology	0	7,4	0,0	0,1	0,3
Electronic Journal of Biotechnology	1	5,6	0,1	0,0	0,1
Food Microbiology	0	4,0	0,0	0,3	0,7
Genes & Genetic Systems	1	4,8	0,0	0,0	0,1
Genes, Brain and Behavior	0	1,9	0,2	0,4	0,8
Genetics and Molecular Biology	1	4,3	0,0	0,1	0,1
Genome	0	4,4	0,0	0,3	0,6
IEEE Transactions on Information Technology in Biomedicine	0	1,4	0,0	0,3	0,7
Indian Journal of Biochemistry and Biophysics	0	4,4	0,1	0,0	0,0
International Journal of Molecular Sciences	0	1,8	0,0	0,1	0,2
International Microbiology	1	5,6	0,4	0,3	0,9
Journal of Animal Ecology	0	1,6	0,0	0,3	0,8
Journal of Basic Microbiology	0	5,6	0,0	0,1	0,3
Journal of Biochemistry and Molecular Biology	1	4,2	0,1	0,1	0,1
Journal of Biomedicine and Biotechnology	1	2,9	0,6	0,5	0,9
Journal of Chemical Technology and Biotechnology	0	4,0	0,0	0,1	0,5
Journal of Clinical Microbiology	0	5,4	0,0	0,3	0,7
Journal of Genetics	1	3,7	0,0	0,0	0,3
Journal of Lipid Research	1	3,4	0,1	0,5	0,8
Journal of Microbiology and Biotechnology	0	7,5	0,0	0,1	0,1
Journal of Molecular Catalysis B: Enzymatic	0	6,6	0,0	0,1	0,3
Journal of Plant Biochemistry and Biotechnology	0	5,5	0,0	0,1	0,1
Journal of Proteome Research	0	4,8	0,0	0,4	0,8
Korean Journal of Genetics	0	6,8	0,0	0,0	0,1
Laboratory Animals	0	4,4	0,1	0,2	0,8
Microbiology and Immunology	1	4,1	0,0	0,1	0,1
Microbiology and Molecular Biology Reviews	0	2,2	1,0	0,4	0,8
Molecular and Cellular Biology	0	3,4	0,0	0,1	0,4

Molecular Biology	0	5,1	0,1	0,0	0,1
Mutagenesis	0	3,3	0,1	0,1	0,8
Nature Biotechnology	0	6,3	0,0	0,6	0,9
North American Journal of Aquaculture	0	2,1	0,0	0,9	1,0
Nucleic Acids Research	1	5,3	0,0	0,4	0,8
Pathobiology	0	1,2	0,1	0,1	0,5
Process Biochemistry	0	6,2	0,0	0,1	0,3
Raffles Bulletin of Zoology	1	3,3	0,0	0,2	0,3
Russian Journal of Genetics	0	4,2	0,1	0,0	0,1
Steroids	0	3,6	0,0	0,2	0,6
Trends in Biotechnology	0	4,3	0,7	0,3	0,8
Trends in Microbiology	0	4,1	0,9	0,4	0,9
Zoosystema	1	2,0	0,1	0,1	0,5

Appendix 2: Mathematics journals included in the study

	OA	Dependency on this data sample	Share of reviews	Share of authors from North America	Share of authors from Western countries
Applied Mathematics and Computation	0	3,6	0,0	0,1	0,1
Annales Academiae Scientiarum Fennicae	1	7,3	0,0	0,2	0,6
Annals of Mathematics	1	7,3	0,0	0,5	0,4
Bulletin of the American Mathematical Society	1	3,9	0,1	0,6	0,3
Communications in Algebra	0	8,0	0,0	0,2	0,3
Discrete Mathematics	0	6,5	0,0	0,3	0,3
Duke Mathematical Journal	0	7,3	0,0	0,4	0,5
Electronic Communications in Probability	1	1,2	0,0	0,3	0,5
Electronic Research Announcements in Mathematical Sciences	1	2,7	0,0	0,4	0,3
Electronic Transactions on Numerical Analysis	1	2,1	0,0	0,4	0,4
European Journal of Applied Mathematics	0	2,6	0,0	0,1	0,5
Forum Mathematicum	0	6,7	0,0	0,3	0,5
Houston Journal of Mathematics	0	7,3	0,0	0,4	0,2
Izvestiya Mathematics	0	4,5	0,0	0,0	0,1
Journal of Nonlinear Mathematical Physics	1	2,9	0,0	0,1	0,3
Journal of the American Mathematical Society	0	7,1	0,0	0,6	0,3
Journal of the London Mathematical Society	0	7,2	0,0	0,2	0,5
Mathematical Logic Quarterly	0	3,6	0,0	0,2	0,4
Mathematical Problems in Engineering	1	1,6	0,0	0,1	0,1
Mathematical Research Letters	0	7,2	0,0	0,5	0,3
Mathematical Social Sciences	0	3,1	0,0	0,2	0,5
NODEA - Nonlinear Differential Equations and Applications	0	2,9	0,0	0,0	0,7
Proceedings of the Japan Academy. Series A.	1	5,1	0,1	0,0	0,1
Quarterly Journal of Mathematics	0	8,1	0,0	0,2	0,5
Studies in Applied Mathematics	0	3,4	0,0	0,4	0,3

Appendix 3: Pharmacy and pharmaceutical journals included in the study

	OA	Dependency on this data sample	Share of reviews	Share of authors from North America	Share of authors from Western countries
AAPS Journal	1	3,3	0,6	0,9	0,1
AAPS PharmSciTech	1	8,6	0,0	0,3	0,2
American Journal of Pharmaceutical Education	1	4,5	0,0	0,9	0,0
Biological & Pharmaceutical Bulletin	1	3,8	0,0	0,0	0,0
British Journal of Clinical Pharmacology	0	4,5	0,0	0,1	0,3
Chemical & Pharmaceutical Bulletin	1	5,3	0,0	0,0	0,1
Clinical and Experimental Pharmacology & Physiology	0	1,9	0,1	0,2	0,1
Current Pharmaceutical Biotechnology	0	0,5	0,9	0,3	0,4
Formulary	0	1,2	0,0	0,9	0,0
Journal of Pharmaceutical and Biomedical Analysis	0	4,8	0,0	0,1	0,3
Journal of Pharmacology and Experimental Therapeutics	0	4,1	0,0	0,5	0,3
Journal of Pharmaceutical Sciences	0	8,0	0,0	0,5	0,3
Journal of Pharmacy and Pharmaceutical Sciences	1	5,0	0,2	0,4	0,1
Journal of Physiology and Pharmacology	1	4,4	0,0	0,1	0,2
Molecular Pharmacology	0	3,7	0,0	0,5	0,3
Pharmaceutical Biology	0	3,1	0,0	0,1	0,1
Pharmaceutical Research	0	5,8	0,1	0,4	0,3
Pharmacological Reports	1	3,4	0,1	0,1	0,1
Pharmacology	0	3,5	0,0	0,1	0,5
Trends in Pharmacological Sciences	0	2,6	0,7	0,3	0,5

10. Attracted to open access journals:

A bibliometric author analysis in the field of biology

Tove Faber Frandsen[14]

Royal School of Library and Information Science
Copenhagen S. Denmark
tff@db.dk

Purpose - Scholars from developing countries have limited access to research publications due to expensive subscription costs. However, the open access movement is challenging the constraint to access. Consequently, researchers in developing countries are often mentioned as major recipients of the benefits when advocating open access (OA). One of the implications of that argument is that authors from developing countries are more likely to perceive open access positively than authors from developed countries. The present study is an investigation of the use of open access by researchers from developing countries and is thus a supplement to the existing author surveys and interviews.

Design/methodology/approach – Bibliometric analyses of both publishing behaviour and citing behaviour in relations to OA publishing provide evidence of the impact of open access on developing countries.

Findings – The results of the multivariate linear regression show that open access journals are not characterised by a different composition of authors than the traditional toll access journals. Furthermore, the results show that authors from developing countries do not cite open access more than authors from developed countries.

Originality/value – This paper argues that authors from developing countries are not more attracted to open access than authors from developed countries.

Keywords – open access, bibliometrics, author analysis, developing countries.

Article Type: Research paper

---

[14] The author wishes to thank Jesper W. Schneider for valuable comments and suggestions.

> "The advent of Open Access strategies […] has the potential to revolutionize access to essential research" (Kirsop & Chan, 2005).

Introduction

The poorest countries [15] will benefit the most from OA initiatives, although some will have a greater impact than others (Kirsop & Chan, 2005). Chan and Kirsop (2001) list several advantages for developing countries: (a) free access to research information from developed countries; (b) researchers in developing countries can archive their own research and thus make it available to researchers in developed countries; (c) local research can be distributed.

Developing countries are often mentioned when describing the advantages of open access: "One of the great beneficiaries of open access may be users in developing countries, where there are currently some universities with no journal subscriptions at all" (Wikipedia.org/wiki/open_access). Along the same lines Xuemao and Chang (2006) state that "OA will benefit the developing countries in the two-way scholarly communication and accelerate the development of science accordingly." Furthermore, Nicholas, Huntingdon and Jamali (2007: 877) makes the prediction about a journal (Nucleic Acids Research) moving to an OA model "(that) it would be assumed that there is some further growth in the pipeline, predominantly from second and third world users".

The open access movement consists of various objectives but essential for the movement is the ambition of distributing peer-reviewed (journal) literature freely

[15] "Poor countries" is one of many terms used to refer to "less developed countries", "developing countries" (see e.g. Chan & Kirsop, 2001 and Kirsop & Chan, 2005) and nations of the "South" (see e.g. Lor & Britz, 2005). Another term is "peripheral countries", although it should be noted that that term is not necessarily used as an equivalent to the aforementioned terms. Peripheral countries in science can be used to describe "not-central countries" in terms of language (Bordons, Fernandez, & Gomez, 2002; Velho, 1986). However, the term can also be used to describe countries "whose economic potential is small and where financial support of science is relatively small" (Puliselic & Petrak, 2006).

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[16]. Both strategies depend greatly on authors being willing to support open access by making their own work available either through self-archiving or open access publishing.

One of the implications of viewing authors from developing countries as great beneficiaries of open access is that such authors are more likely to perceive OA positively than authors from developed countries as it enables them to access research that they otherwise would have had limited access to due to expensive subscription costs. Correspondingly, authors from developed countries should perceive OA less positively than authors from developing countries as they already have access to the necessary research publications within their field.

However, this hypothesis is somewhat contradicted by the results of Swan and Brown (2005) as they find the perception of open access to be tied to a number of other factors than the geographic location of the author.[17] One of the results from the survey is that the principle of free access is the strongest imperative for publishing in an open access journal. Furthermore, authors publishing in an open access journal consider it to have a larger audience than toll-access journals, to publish more rapidly and to be prestigious in their field. On the other hand, authors who have not (yet) published in an open access journal state that it is due to being unfamiliar with open access journals in their field and being unable to identify a suitable open access journal to publish in. Perception of open access journals is closely related to having published in an open

[16] 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).

[17] The response rate (about 5 per cent) in the survey by Swan and Brown (2005) must be taken into account as such small samples may not necessarily be generalisable.

access journal or not. Swan and Brown (2005) analyse geographical relations on the basis of the region of the author. However, the geographical origin is only included in a very limited amount of their analyses, and the number of respondents from the developing countries are small (e.g. 52 respondents from African countries and 39 from China). Finally, analysing the questions by geographic origin does not reveal a clear pattern.

Furthermore, as stated by Papin-Ramcharan and Dawe (2006), developing countries may not be able to fully benefit from the advantages of OA:

> "Although there are the obvious benefits to developing country researchers of free access to the scholarly literature, there are many hindrances to such researchers fully benefiting from this largesse." (Papin-Ramcharan & Dawe, 2006: 24)

Papin-Ramcharan and Dawe (2006) move on to claim that developing countries are struggling with technical, financial, human and infrastructural limitations making the scholars in these countries unable to access the research publications of others and disseminate their own research results the way it is intended by the open access movement. However, the specific issues of developing countries are easily overlooked when focusing on the many difficulties of achieving open access (see e.g. Björk, 2004).

The purpose of the present study is to analyse the behaviour of authors in relation to OA publishing. More specifically, the aim is to examine whether the behaviour of authors can be tied to author characteristics (in terms of geographic location) or journal characteristics (being familiar with the open access principle). The main research question can be elaborated into the following sub-questions:

- Are authors from developing countries more attracted to publishing in OA journals? Are OA journals thus characterised by a greater share of authors from developing countries than traditional subscription based journals?
- Do authors from developed and developing countries cite differently in OA journals than authors of same nationalities publishing in NOA journals?

Related research

This study draws primarily on existing studies of author analyses in relation to OA and analyses of the degree of internationality.

The existing literature provides numerous author analyses in relation to OA and several of them indicate various author reservations. Author reservations can be seen falling in three categories: (a) willingness to pay, (b) perception of quality and (c) awareness of OA. Willingness to pay for OA is a complicated issue as authors are willing to pay although not the full costs. Cozzarelli, Fulton and Sullenberger (2004) conduct a survey of authors having their paper accepted for publishing in The Proceedings of the National Academy of Sciences USA (PNAS) in a two month period in 2003. Although restricted to a very small data sample their analysis shows that almost 50 per cent of the authors would be willing to pay a surcharge to make their PNAS article open access. However, almost 80 per cent of those willing to pay would only agree to a surcharge of $500 which is in contrast to the estimated costs in 2004 of $2500 (Wellcome Trust, 2004). Richardson and Saxby (2004) investigate the willingness of authors to pay an author charge of $500 per article in the journal Nucleic Acids Research and attain agreement from 90 per cent of the papers in an issue in January 2004. The surveys by Rowlands, Nicholas, and Huntingdon (2004a,b) and Nicholas, Huntingdon and Rowlands (2005) reveal discipline variances in the willingness to pay author charges.[18]

Secondly, the perceived quality is a crucial issue in the development of OA. Hess et al. (2007) study the perception of OA publishing based on 688 responses to an online survey. They find very positive attitudes towards the OA principle in general. However, at the same time authors are reluctant to publish their own work in OA journals. Pelizzari (2003) surveys the staff at two faculties of a single university, but the sample size is too limited to be able to make any solid conclusions. Furthermore, a survey by Schroter and Tite (2006) including 468 respondents (a response rate of 42 per cent) reveals that 66 per cent of authors would prefer publishing in a journal that is not an open access author-pays journal. Furthermore, Schroter, Tite and Smith (2005) interview 28 authors, and although the results can hardly be generalised, they conclude

[18] It should be noted that the response rate is very low (less than 5 per cent).

that authors perceive journal quality to be more significant than any potential author charges when deciding on a journal to submit a paper to.

Thirdly, many authors are not aware of open access (Rowlands, Nicholas & Huntingdon, 2004a,b; Nicholas, Huntingdon & Rowlands, 2005). Schroter, Tite and Smith (2005) find that authors are increasingly aware of open access although few have published in an OA journal.

Summing up the existing author analyses, it is fair to say that they are mostly limited to experiments with heavily subsidised author charges, low response rates and limited samples. Furthermore, they are all based on either surveys or interviews. There seems to be a lack of investigations on a larger scale focusing less on the opinions of authors and more on the actual publishing and citing behaviour of authors. Large scale investigations are possible using bibliometrics, however, the existing bibliometric studies have focused modestly on OA and author behaviour.

The degree of internationality has been the subject of several analyses. The term "international" lacks a clear definition leading to a highly ambiguous term according to Buela-Casal et al. (2006). Their aim is to construct a set of criteria designed to measure the degree of internationality and identify four core criteria. The first is the multinational collaboration patterns (the share of articles written by at least two authors affiliated to institutions from two different countries). The remaining criteria are the multinational distribution of editorial board members, associate editors and authors. Using these criteria or another set of criteria the analyses can be made on various levels. One approach is on the level of countries (e.g. Glänzel & Schubert, 2005; Schubert & Glänzel, 2006) and another is on the level of disciplines (e.g. Abt, 2007; Jappe, 2007). A subset of the analyses of internationality is focused on developing countries and the centre of attention is typically on the collaboration patterns of one or a few developing countries (e.g. Kim, 2005; Royle et al., 2007), and in some cases the studies are also delineated to only one or a few disciplines (e.g. Kim, 2006).

As can be seen from this review of the existing literature, to make definitive conclusions about the potentials of open access for developing countries, there is a need for more studies in this research area, especially those analysing developing countries at a larger

scale and investigate the actual publishing and citing behaviour of authors from those countries.

Data

This paper is based on two analyses. Although the first is a publication analysis and the second is a citation analysis, the starting points of the two analyses are the same. The dataset collected for the analyses includes journals from the discipline of biology using the classification scheme of Ulrich's Periodicals Directory™ (Ulrich's). Biology is a field characterised by a large number of OA journals making it a suitable field for this study. The dataset consists of a selection of 150 journals that are either general biology journals or belong to one or several of the following sub-disciplines: entomology, zoology, microbiology, biochemistry, genetics, and biotechnology.

The classification scheme used by Ulrich's is only one scheme among many possible as countless different classification systems can be used to divide a discipline. First and foremost the discipline needs to be determined which may not necessarily be an easy task. Levine (1965) uses the terms bio sciences, life sciences, human biology and biology. Furthermore, Levine (1965: 346) states that "(t)he particular system used is determined by preference and proposed use." According to Dullemeijer (1980) the principle of a division in biology can be made on objects and aspects. He stresses that "(d)ividing biology into subdisciplines is not merely an amusing game, but the basis for management in a science policy and a necessity for theoretical biology" (Dullemeijer, 1980: 87). It must be noted that the division used in this study is made on journal level and not on the more detailed level of articles. The unit of analysis is a group of articles and it is thus just as difficult to determine their sub-discipline as a whole. Consequently, the subject division is made on journal level and for that purpose the scheme of Ulrich's is ideal.

The set contains 150 journals, however, some journals lacked information for certain variables and were excluded from one or both of the models (resulting in 130 journals in the first analysis and 136 in the second). A list of the included journals is available in appendix 1. Furthermore, it should be noted that the analysis was performed on the

basis of data from 2006 as that is the most recent year completed in the citation indexes at the time of the data gathering process.

The unit of analysis was the publications of the 150 journals. Initially, the publications of all journals were divided into three groups in terms of the geographic location of the author(s): (a) publications written by author(s) all located in developed countries, (b) publications written by author(s) from developing countries, and (c) publications written by author(s) from developed countries as well as author(s) from developing countries. The division of countries into two categories (developing and developed) is made on the basis of the country classification by The World Bank which is based on gross national income (GNI) per capita.[19] Based on its GNI per capita, every economy is classified as low income, middle income (subdivided into lower middle and upper middle), or high income (www.worldbank.org). In this study the developing countries are defined as those belonging to either the group of low income countries or the group of lower middle income countries.

It should be noted that some publications lack information in the geographic location-field and they are thus excluded from the analysis. However, the lack of information in some publications is not directly related to the research questions of this study and thus we can expect the excluded publications to be distributed randomly across the dataset.

In order to be able to analyse the publishing and citing behaviour of authors some control variables were needed as means to ensure that underlying factors are not distorting the results. If, say, journals from developing countries are more likely to be OA than journals from developed countries and if authors from developing countries are more likely to publish their work in journals from developing countries, then failing to control for the geographic location of the journal would lead to a bias in the estimates, as it would appear as if developing authors are more likely to publish OA, even though it is in fact not the case, since all they do is to publish their work locally. Indeed, the analysis below confirms that this is in fact the case.

[19] There are a number of problems related to this definition of poor countries. One issue is the skewed income distribution that implies that a country with a relatively high per-capita GNI could have a majority of citizens with relatively low levels of income, due to concentration of wealth in the hands of a small fraction of the population. Although, the definition is arguable, it is applicable for operational and analytical purposes.

Information about the journals included in the study was collected using Ulrich's, the Journal Citation Reports (JCR) and the ISI (or the Institute for Scientific Information) citation database called the Science Citation Index (SCI) which can be searched in what is currently known as Web of Science. As already mentioned the classification scheme of Ulrich's has been used to divide the journals into sub-disciplines. Each journal can belong to more than one sub-discipline depending on the classification in Ulrich's (the journals belong to a maximum of three sub-disciplines). Furthermore, information on the geographic location of the journals was available through Ulrich's. However, it should be noted that the geographic location listed by Ulrich's may not necessarily depict the whole picture as some journals may be listed with the location of the publisher although the editorial board members and associate editors are located in a completely different parts of the world. However, although using the geographic location available through Ulrich's is not unblemished it can be a valuable variable to include as a journal located in a developing country is likely to have a higher share of authors from that country. Ulrich's was also used to determine the OA status of the journal as the website provides a link to the freely available full text of the journal. JCR can provide information on the journal impact factor (JIF) of the journal. Although highly disputed JIF is probably the central quantitative indicator for measuring journal quality.[20] The most recent edition (at the time of the data gathering process) of the JCR was used to collect the data (the 2005 edition). The SCI has been used to collect data on the share of publications in the journal that is not written in English. On the one hand, journals written in Hindi, Spanish or Russian are expected to have a smaller audience than journals written in English. The audience is smaller simply because more scholars are able to read English than e.g. Hindi. A journal in Hindi would primarily attract Indian authors and thus the variable was added to control for that possible bias on the

[20] The idea of an impact factor first emerged in 1955 (Garfield, 1955) but it was not expected to become the subject of widespread controversy. The inventor himself describes the impact factor as a mixed blessing (Garfield, 1999). The impact factor was created in order to be able to compare journals regardless of their size in order to perform the journal selection for SCI. The intent of the IF was to provide an objective method for quantifying published research output, rank and evaluate journals (Garfield, 1996). According to the ISI JIF is a measure of past usage of journals and is an indicator of the effectiveness of a journal within the wider context of the international research community (O'Neill, 2000). JIF rapidly became accepted and widely used because of its comprehensibility, robustness and availability (Glänzel & Moed, 2002). Although extremely critical, Bordons, Fernandez & Gomez (2002) emphasize that JIF give a visible form to the invisible hierarchy of scientific journals that is tacitly accepted by scientists.

results. Furthermore, the share of reviews was determined using SCI. The variable was added to capture the possible effect of journals publishing many reviews may cite OA journals less. One explanation is a reluctance to accept OA articles as a part of state-of-the-art within a discipline. Another is that because of the higher average age of references in review papers (see e.g. Price, 1965) and the fact that many OA journals have only existed for 2 to 3 years, review papers have a lower share of references to OA journals.

Finally, the set of cited OA journals needs to be constructed for the analysis of citing behavior. For that purpose a list of active OA journals was essential, however, as no comprehensive directory of OA journals exists the lists of biology OA journals in Ulrich's and the Directory of Open Access Journals (DOAJ) was merged and duplicates were removed (data gathering took place in August 2007). Combining the two lists yielded 427 OA journals (279 on the list from Ulrich's and 274 on the list from DOAJ). However, there was an overlap of 126 journals (30 per cent) resulting in 301 unique OA journals. It is evident that the size of the overlap is not impressive and a combination of the two lists provides a more comprehensive list.

Analysis of the publishing behavior

The dependent variable in this analysis is the share of publications in the journal written exclusively by authors from developing countries.[21] This data was collected using the SCI searched. An example could be a journal comprising of 35 publications in 2006 made up by 25 publications by authors from developed countries, 7 by authors from both developed and developing countries and finally 3 by authors from developing countries. In this case the percentage of publications in the journal written exclusively by authors from developing countries is 8.6. A minimum of 10 publications per year for each journal was set as the inclusion of journals with fewer publications led to very unstable figures. The model used for this analysis includes the following independent variables: (a) geographic location of journal; (b) share of publications not written in

[21] The analysis could be done on the level of authors. However, an analysis on that level requires a fractional scheme that attributes to that paper a fraction of countries or regions. An integer scheme, however, assigns the paper integrally or wholly to countries or regions (see e.g. Moed, 2005: 273-274). In this case the scheme is integral, although a paper can only be assigned to one of the two regions (developing or developed) or in a category containing publications which have authors from both regions.

English; (c) share of reviews; (d) OA status (dummy variable[22]); (e) sub-disciplines (dummy variables); and (f) JCR JIF. An overview of the dependent and independent variables is available in Table 1.

Variable	Range	Mean	Std. Dev.
Dependent variable			
Percentage of authors from developing countries	0 – 86	12.33	16.71
Independent variables			
JIF	0.027 – 22.74	2.79	3.69
OA	0 - 1	.26	.44
Sub-discipline: Entomology	0 - 1	.16	.37
Sub-discipline: Zoology	0 - 1	.16	.37
Sub-discipline: Microbiology	0 - 1	.15	.36
Sub-discipline: Biochemistry	0 - 1	.26	.44
Sub-discipline: Genetics	0 - 1	.15	.36
Sub-discipline: Biotechnology	0 – 1	.16	.37
Geographic location of journal: developing country	0 – 1	.25	.44
Share of reviews	0 – 1	.13	.26
Share of articles not in English	0 – 0.97	.02	.13

Table. 1. The range of the dependent and independent variables included in the multivariate linear regression regarding publishing behaviour.

The present model is not characterized by dependent variables being highly inter-correlated with each other. The highest inter-correlation is found between the following variables: JIF is related to the share of reviews (r = .490). Journals containing a high proportion of review articles tend to have higher IFs than other journals which Garfield (1996) also notes. However, adding the statistically significant correlating variables to the model as interaction variables did not contribute significantly to the explanation of

[22] A variable that takes on the values 0 or 1 and is used to describe the effects of the different levels of a qualitative independent variable in a regression model (Bowerman & O'Connel, 2007)

the dependent variable and were consequently excluded from the model. An overview of the inter-correlations is available in appendix 2.

Analysis of the citing behavior

The dependent variable in this analysis is the share of references in the journal to OA journals. This data was also collected using the SCI searched in Web of Science. The method of calculating the share of references to OA journals is not entirely the same for OA and NOA journals. Journal self-citations make up a considerable amount of citations both in terms of self-citing and self-cited rate (Frandsen, 2007; Tsay, 2006). Including journal self-citations would lead to biased results as they would affect the calculation of OA shares for OA journals more than NOA journals. But as stressed by White (2001) self-citations are not an insurmountable difficulty as they can be excluded from the analyses. In this case self-citations only pose a problem for the calculation of references to OA journals in OA journals and consequently the self-citations were omitted from these calculations. The share of references to OA journals from OA journals is calculated as follows:

$$\text{Share of references from OA to OA} = \frac{\text{Number of references to OA journals} - \text{self-citations}}{\text{Total number of references} - \text{self-citations}}$$

The calculation of the share of references to OA journals from NOA journals is slightly different as the numerator does not include self-citations. The share is calculated as follows:

$$\text{Share of references from NOA to OA} = \frac{\text{Number of references to OA journals}}{\text{Total number of references} - \text{self-citations}}$$

A minimum of 300 references per year for each of the three author groups was set as the calculation of shares on the basis of fewer references led to very unstable figures. The model used for this analysis includes the following independent variables: (a) geographic location of journal; (b) share of publications not written in English; (c) share of reviews; (d) OA status (dummy variable); (e) sub-disciplines (dummy variables); and

(f) JCR JIF. Table 2 provides an overview of the included variables. It should be noted that the ranges and means differ from the dataset constructed to analyse publishing behaviour as the same journal can be represented by the three different groups of authors and some of the journals excluded in the former analysis can be included in the latter.

Variable	Range	Mean	Std. Dev.
Dependent variable			
Percentage of references to the set of OA journals	0 – 11.31	1.67	1.66
Independent variables			
Author(s) from developing countries	0 - 1	.20	.40
Author(s) from both developed and developing countries	0 - 1	.15	.36
JIF	0.027 – 22.74	2.77	3.28
OA	0 - 1	.26	.44
Sub-discipline: Entomology	0 - 1	.13	.34
Sub-discipline: Zoology	0 - 1	.13	.33
Sub-discipline: Microbiology	0 - 1	.17	.38
Sub-discipline: Biochemistry	0 - 1	.30	.46
Sub-discipline: Genetics	0 - 1	.15	.36
Sub-discipline: Biotechnology	0 – 1	.18	.38
Geographic location of journal: developing country	0 – 1	.18	.38
Share of reviews	0 – 1	.12	.25
Share of articles not in English	0 – 0.97	.01	.10

Table. 2. The range of the dependent and independent variables included in the multivariate linear regression regarding citing behaviour.

Highly inter-correlated independent variables do not pose a greater problem to this model than the previous one. The highest inter-correlation is found between the following variables: JIF is related to the share of reviews (r = .445). However, adding an interaction variable for these and the other statistical significant inter-correlated variables did not contribute significantly to the explanation of the dependent variable

and the interaction variables were consequently excluded from the model. An overview of the inter-correlations is available in appendix 3.

The analyses below consist of multivariate linear regression analyses of the statistical relations between the dependent and the independent variables.

Results

The results of this study are presented in two sections: (a) publishing behaviour; and (b) citing behaviour.

Publishing behavior

The unit of analysis (and thus number of observations) for this analysis consists of 130 (journals). The share of authors from developing countries is 12 per cent in the NOA journals and 14 per cent in the OA journals. The higher level of authors from developing countries in OA journals does not change by including the publications co-authored with authors from developed countries although the difference reduces slightly. However, as already mentioned a number of factors can be influencing the share of authors from developing countries in journals and a multivariate linear regression makes it possible to control for the effects of those factors. Table 3 shows the results of the multivariate linear regression.

Variable	Coefficients	P-value
Intercept	12.99	.00
JIF	**-1.21**	**.01**
OA	-4.51	.16
Sub-discipline: Entomology	-2.69	.54
Sub-discipline: Zoology	-2.72	.53
Sub-discipline: Microbiology	-5.33	.18
Sub-discipline: Biochemistry	5.25	.11
Sub-discipline: Genetics	-3.70	.34
Sub-discipline: Biotechnology	6.14	.12
Geographic location of journal: developing country	**15.26**	**.00**
Share of reviews	3.78	.52
Share of articles not in English	**-25.71**	**.02**
R squared	.311	
# of observations	130	

Table 3. Multivariate linear regression analysis. Dependent variable: share of authors from developing countries excluding authors co-authoring with authors from developed countries. Variables statistically significant at the .05 level are marked with bold.

The most important number in Table 3 is in fact a variable that is not statistically significant. The third row of the table showing the OA-variable has a p-value of .16. This implies that OA journals do not differ from NOA journals with respect to the share of publications written by authors from developing countries. The difference in the percentage of authors from developing countries in NOA and OA journals is consequently not related to the OA status of the journal but (at least to some extent) other variables.

Turning now to the coefficients of the variables that enter the estimation significantly Table 6 shows that there are three such variables at the .05 level: (a) JIF; (b) geographic location of the journal being a developing country; and (c) share of articles not in English. The values of the coefficients to these variables can be interpreted as follows: (a) an increase by one unit in the JCR impact factor leads to a decrease in the share of publications written by authors from developing countries of 1.2 percentage points; (b) a geographic location of the journal in a developing country leads to an increase of 15

percentage points; and (c) for every increase in the share of articles not in English of 10 per cent, the share of publications written by authors from developing countries decreases by 2.5 percentage points. JIF is related to the share of authors from developing countries. However, it must be stressed that causation cannot be determined and it could be the share of authors from developing countries affecting the JIF. Furthermore, the coefficient of -1.21 implies that a decrease in the share of publications written by authors from developing countries of 1.2 percentage points is statistically associated to an increase in JIF by 1 unit. Given that the overall average of the share of publications by developing authors is around 12-13 per cent quite a substantial degree of the variation can thus be explained by differences in JIF across the journals in the sample. The negative effect of the non-English language variable on the share of authors from developing countries is much greater (and less complicated to change). If a journal with all their publications written in a non-English language changes the language to English it implies an increase in the share of authors from developing countries by 25.71 percentage points. To a large extent the effect is caused by the specific languages of the journals included in this study. Russian is one of the included languages that probably will not attract many authors from developing countries (Russia is not characterised as a developing country according to the country groups by The World Bank). Non-English languages are associated with a more national profile in the author distribution and changing the language to English does attract more foreign authors, including authors from developing countries (Dinkel et al., 2004; Puliselic & Petrak, 2006).

A result of the calculations performed with programs such as SPSS is the R-square, which is used as indicators of the ‘‘goodness of fit.’’ Generally speaking, an R-square (percentage of total variance) that is close to 1 is indications that the data fit the model well. The R-square values of this model and the following model show that about 30 per cent of the variance is explained by the selected variables. Higher R-square values would have indicated a better fitted model, although, this level of R-squares is not atypical in analyses of this character. Multivariate analysis dealing with social science data typically result in R-squares ranging from 0.1 to 0.5 (Lattin; Carroll & Green, 2003). Data on the behaviour of scholars at the level of single papers or authors tend to be largely dispersed with values widely scattered. Schrum and Campion (2000) analyse the degree of isolation of scientists in developing countries and their multiple regression analyses have R-squares of 0.17-0.21. Nemeth and Goncalo (2003) study the citation

impact of collaborations between authors at different universities or locations and the R-squares are 0.067 and 0.068. A much better fit is achieved at the level of countries (Tijssen & Leeuwen, 2006) or journals (Frandsen, 2007). Although, the models in this study do not fit the data excellently there are variables that can explain about 30 per cent of the variance.

Citing behavior

The unit of analysis (and thus number of observations) for this analysis consists 200 groups of publications. There are publications by author(s) from developed countries, author(s) from developing countries and author(s) from both developing and developed countries. In principle the dataset could consist of 150 * 3 observations, however, as some journals had no or too few publications by one or more of the three author groups the dataset is reduced to 200 observations.

The lowest average shares of references to OA journals are found in publications by authors from developed countries as the average is 1.38 per cent. Slightly higher average shares are found in publications written by authors from developing countries as the average share is 1.59 per cent, and finally the publications from both developed and developing countries have the highest average share of 2.46 per cent. Given these crude numbers it thus appears that developing authors use OA less, and not more, than authors from developed countries. However, in order to be able to understand the factors at work, a multivariate linear regression model is applied to control for a number of factors that may influence the citing behaviour of the authors. Table 4 presents an overview of the results.

Variable	Coefficients	P-value
Intercept	1.13	.00
Author(s) from developing countries	-.09	.75
Author(s) from both developed and developing countries	**.93**	**.00**
JIF	**-.11**	**.01**
OA	**.56**	**.02**
Sub-discipline: Entomology	**-.73**	**.05**
Sub-discipline: Zoology	.37	.32
Sub-discipline: Microbiology	-.15	.63
Sub-discipline: Biochemistry	-.42	.10
Sub-discipline: Genetics	**1.77**	**.00**
Sub-discipline: Biotechnology	-.07	.82
Geographic location of journal: developing country	.13	.68
Share of reviews	**-.95**	**.04**
Share of articles not in English	1.86	.08
R squared	.306	
# of observations	200	

Table 4. Multivariate linear regression analysis. Dependent variable: share of references to OA journals. Variables statistically significant at the .05 level are marked with bold.

Of special interest to this study Table 4 clarifies that the share of references to OA journals is not different between authors from developed and developing countries. Consequently, authors from developing countries do not cite OA significantly more (or less) than authors from developed countries. However, the OA status of the journal is related to the share of references to the set of OA journals implying that authors publishing in OA cite other OA journals more than authors publishing in NOA journals. Going through the model, Table 4 shows that it has six coefficients that are statistically significant at the .05 level: (a) author(s) from both developed and developing countries; (b) JIF; (c) OA; (d) the sub-discipline entomology; (e) the sub-discipline genetics; and (f) the share of reviews in the journal. The values of these coefficients can be interpreted as follows: (a) publications by author(s) from both developed and developing countries have a share of references to OA journals of .93 percentage points higher; (b) an

increase by one unit in the JCR impact factor corresponds to a decrease in the share of references to OA journals by .11 percentage points; (c) OA journals have shares of references to OA journals of .56 percentage points higher; (d), the sub-discipline entomology have shares of references to OA publications .73 percentage points lower; the sub-discipline genetics have shares of references to OA publications 1.77 percentage points higher; and (f) an increase in the share of reviews in the citing journal leads to a decrease in the share of references to OA journals of .95 percentage points. Finally, it should be noted that the variable describing the share of articles not in English is significant at the 0.1 level indicating that journals publishing many articles not in English may cite OA journals more.

Discussion

First of all, the sub-disciplinary differences in the share of references to OA journals should be noted. It is well established in the existing literature that the use of Internet based communication varies considerably from field to field. These differences have been affirmed in the existing literature on the level of disciplines (e.g. Barjak, Li & Thelwall, 2007; Kling & McKim, 2000; Törmä & Vakkari, 2004)) as well as sub-disciplines (e.g. Frandsen & Nicolaisen, 2007; Fry, 2006; Sotudeh & Horri, 2007). In the present analysis not all the sub-disciplines are statistically significant variables contributing to the understanding of the dependent variable. However, as stated by Kling and McKim (2000: 1307) there are great differences "from one field (or closely related set of fields) to another." Some of the sub-discipline specific journals are thus not differing from the general biology journals whereas other sub-discipline specific journals cite OA journals either less or more than the general biology journals.

The results of this study indicate that having controlled for other factors authors from developing countries are not more attracted to OA journals than authors from developed countries. Furthermore, the results indicate that authors from developing countries do not cite OA journals more than authors from developed countries. Authors from developed and developing countries thus have similar publishing and citing behaviour which could be a reflection of the fact that the incentives in terms of publishing and citing being equivalent. Authors from developing countries are competing for attention just as authors from developed countries. Attention is a scarce resource in science and an object of competition (Klamer & van Dalen, 2002). An author can gain attention

through reputation, the journals the author publishes in and the citations he or she receives (van Dalen & Henkens, 2005). Reputation is, however, closely related to the latter two ways of gaining attention. Appointments, promotions, and grant applications are strongly influenced by publication records (Daniel, 2005) and citation records (Lewison et al. (1995; Saiz-Salinas, 1996). The distribution of publications and citations among individuals, departments, institutions and countries is thus an increasingly important feature of contemporary science. The increased focus on evaluation based on publication and citation analyses has according to Kaltenborn and Kuhn (2004) led authors more or less voluntarily to adapt their publication strategy to a maximation of their impact. Authors seek to and are pressured to publish in journals with high JIFs (Maffuli, 1995; Vinkler, 1986). The authors from developing countries compete to get their research published in the same high JIF journals as the authors from developed countries. The survey by Alemna, Chifwepa and Rosenberg (1999) indicates that although authors from African countries use African journals the greater importance of non-African journals to teaching and research remains undisputed.[23]

On the basis of the data included in this study authors cannot be characterised on the basis of being located in either a developed or a developing country. However, it must be stressed that the authors from developing countries in this study may not be representative of the population of authors from these countries. The authors included in this study have been able to get their research published in journals included in the Science Citation Index which could be related to the topic of their research (being of interest outside their own country or region). Tijssen et al. (2006) find that African journals are to a large extent invisible in international databases, and Shrum (1997) finds that many of the characteristics of the scientists from developing countries that are visible in international databases differ from the wider population of scientists from developing countries. Consequently, the international databases depict a picture of the scientists from the developing countries that is not accurately reflecting the population of researchers or domestic productivity.

The results also indicate that OA journals as a publishing medium cannot be characterised as particularly attracting authors who have less access to NOA journal articles and consequently have to depend more on the publications freely available in

[23] In spite of the dominating international journals there are several motives for African scientists not to publish in international journals (Pouris & Richter, 2000).

OA journals. However, OA journals can be characterised by attracting a certain group of authors. Authors publishing in OA journals are citing other OA journals more. The share of references to OA journals (excluding journal self-citations) are .56 percentage points higher than in NOA journals.

Finally, the results show that although authors from developing and developed countries do not differ in terms of citing OA journals, publications by both authors from developed and developing countries differ from the two former groups. Their share of references to OA journals is on average more than 1.5 times greater than equivalent share in publications by authors from developed countries and about 1.8 times greater than the share in publications by authors from developing countries. First of all it must be stressed that this sample of publications is more likely to contain co-authored publications than the other two as this group must consist of two geographic locations as a minimum. International scientific collaboration is an area of great attention in bibliometrics (for recent contributions see e.g. Bookstein, Moed & Yitzahki, 2006a,b; Glänzel & Schubert, 2005; Jappe, 2007 and Ponds, van Oort & Frenken, 2007). Collaboration is typically analysed through co-authorships, although co-authorships are only a partial indicator of collaboration which may not provide the full picture (Katz & Martin, 1997 and Laudel, 2002). Two key issues of interest in international scientific collaboration are the growth and the potential benefits of the collaboration. International collaboration is growing and this growth can be explained by internal as well as external factors to science (Wagner & Leydesdorff, 2005). Moed (2005: 285) provides an example of an analysis of the benefits of international collaboration. Unfortunately, studies on international collaboration including the perspective of distance are few. Collaboration characterized by a long distance between the participating researchers can also be described as remote collaboration (Lorigo & Pellacini, 2007), collaboration with weak ties (Wagner, 2005), collaborators-with-a-difference (Wagner, 2005) or dispersed collaboration (Cummings & Kiesler, 2005). The distance of the collaborators can be defined in a number of ways, e.g. geographical (e.g. Ponds, van Oort & Frenken, 2007 studying scientific collaboration in the Netherlands), disciplinary (e.g. Cummings & Kiesler, 2005 studying multidisciplinary research projects) or financial (the present study using GNI). However, very few studies focus on international collaboration by collaborators-with-a-difference. An essay from 2005 discusses the globalization of science by focusing on collaboration in terms of guests between developing and developed countries (Shrum, 2005: 737). He argues that data collection is extremely

difficult if a representative sample is to be collected in the developing countries. Furthermore, conditions there are fundamentally very different from those in developed countries and thus comparisons are almost impossible to make. An example of the different conditions is provided by Duque et al. (2005: 756) stating that collaboration has become a scientific value in itself. However, they find that collaboration in three African countries is not associated with a general increase in productivity which is the case in developed countries. Consequently, collaboration becomes a paradox for researchers in those countries as their investments in collaboration do not seem to provide them with a fair return. Concluding the discussion of collaborators-with-a-difference it should be noted that further research is needed to determine why the publications by both authors from developed and developing countries differ from publications by authors from just one of those two groups. It could be caused by the lack of a representative sample or the fact that a sample with many co-authored publications is being compared to samples consisting of fewer co-authored publications. However, it could also be caused by inherent characteristics of collaborations by collaborators-with-a-difference.

Conclusion

The results of this study indicate that authors from developing countries are not more attracted to OA journals than authors from developed countries. Authors from developing countries do not publish more in OA journals than authors from developed countries and furthermore, authors from developing countries do not cite OA journals more than authors from developed countries. Consequently, based on this study author behaviour in terms of OA publishing and citing cannot be distinguished on the basis of the author(s) being located in developed or developing country. However, OA journals can be characterised by attracting a certain group of authors as the results show that although authors from developing and developed countries do not differ in terms of citing OA journals, publications by both authors from developed and developing countries differ from the two former groups.

Because the data used in this study does not include other demographic information than the geographic location, it would be interesting to see future research on the topic of attraction to OA include e.g. the age of the author(s). Further, it will be important to test the generalizability of the present findings by (a) investigating the phenomenon not

only in biology fields but also as practiced in other disciplines, (b) investigating other OA models such as self-archived publications as well as open access published publications, and (c) including usage data to analyse reading behaviour as a supplement to publishing and citing behaviour. Finally, content analyses of publications could provide valuable information of qualitative differences citing behaviour.

Literature

Abt, H. A. (2007). The frequencies of multinational papers in various sciences. Scientometrics, 72(1), 105-115.

Alemna, A; Chifwepa, V. and Rosenberg, D. (1999). African Journals: An Evaluation of the Use Made of African Published Journals in African Universities. London: Department for International Development. Education Research Paper No. 36. Available at: http://www.dfid.gov.uk/Pubs/files/africanjournalsedpaper36.pdf

Barjak, F.; Li, X. M. and Thelwall, M. (2007). Which factors explain the Web impact of scientists' personal homepages? Journal of the American Society for Information Science and Technology, 58(2), 87-108, 200-211.

Björk, B. C. (2004). Open access to scientific publications – an analysis of the barriers to change? Information Research, 9(2), paper no. 170. Available at: http://informationr.net/ir/9-2/paper170.html

Bookstein, A.; Moed, H. and Yitzahki, M. (2006a). Measures of international collaboration in scientific literature: Part I. Information Processing & Management, 42(6), 1408-1421.

Bookstein, A.; Moed, H. and Yitzahki, M. (2006b). Measures of international collaboration in scientific literature: Part II. Information Processing & Management, 42(6), 1422-1427.

Bordons, M.; Fernandez, M. T. and Gomez, I. (2002). Advantages and limitation in the use of impact factor measures for the assessment of research performance in a peripheral country. Scientometrics, 53(2), 195-206.

Bowerman, B. L. and O'Connell, R. T. (2007). Business Statistics in Practice. McGraw-Hill/Irwin, 4th edition.

Buela-Casal, G.; Perakakis, P.; Taylor, M. and Checa, P. (2006). Measuring internationality: Reflections and perspectives on academic journals. Scientometrics, 67(1), 45-65.

Chan, L. and Kirsop, B. (2001). Open Archiving Opportunities for Developing Countries: towards equitable distributionof global knowledge. Ariadne, 30. Available at: http://www.ariadne.ac.uk/issue30/oai-chan/intro.html

Cozzarelli, N. R.; Fulton, K. R. and Sullenberger, D. M. (2004). Results of a PNAS author survey on an open access option for publication. Proc Natl Acad Sci 2004;101: 1111. Available at: www.pnas.org/cgi/doi/10.1073/pnas.0307315101.

Cummings, J. N. and Kiesler, S. (2005). Collaborative research across disciplinary and organizational boundaries. Social Studies of Science, 35(5), 703-722.

Daniel, H.-D. (2005). Publications as a measure of scientific advancement and of scientists' productivity. Learned Publishing , 18, 143-148.

Dinkel, A.; Berth, H.; Borkenhagen, A. and Brahler, E. (2004). On raising the international dissemination of German research: Does changing publication language to English attract foreign authors to publish in a German basic psychology research journal? Experimental Psychology , 51,(4), 319–328.

Dullemeijer, P. (1980). Dividing biology into disciplines: Chaos or multiformity? Journal Acta Biotheoretica, 29(2), 87-93.

Duque, R. B.; Ynalvez, M.; Sooryamoorthy, R.; Mbatia, P.; Dzorgbo, D.-B. and Shrum, W. (2005). Collaboration Paradox: Scientific Productivity, the Internet, and Problems of Research in Developing Areas. Social Studies of Science, 35(5): 755–85.

Frandsen, T. F. (2007). Journal self-citations - analysing the JIF mechanism. Journal of Informetrics, 1(1), 47-58.

Frandsen, T. F. & Nicolaisen, J. (2007). Intra-disciplinary differences in database coverage and the consequences for bibliometric research. Submitted to Journal of the American Society for Information Science and Technology.

Fry, J. (2006). Scholarly research and information practices: a domain analytic approach. Information Processing and Management, 42(1), 299-316.

Garfield, E. (1955). Citation Indexes for Science: A New Dimension in Documentation through Association of Ideas. Science, 122(3159), 108-11.

Garfield, E. (1996). How can impact factors be improved? British Medical Journal, 313(7054), 411-413.

Garfield, E. (1999). Journal impact factor: a brief review. Canadian Medical Association Journal, 161(8), 979-80.

Glänzel, W. and Moed, H.F. (2002). Journal impact measures in bibliometric research. Scientometrics, 53(2), 171-193.

Glänzel, W. and Schubert, A. (2005). Domesticity and internationality in co-authorship, references and citations. Scientometrics, 65(3), 323-342.

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

Hess, T.; Wigand, R. T.; Mann, F and von Walter, B. (2007). Management Report 1/2007. Open Access & Science Publishing, Ludwig-Maximilians-Universität Munich, Germany Institute for Information Systems and New Media In cooperation with: University of Arkansas at Little Rock, USA, Department of Information Science.

Available at: http://openaccess-study.com/Hess_Wigand_Mann_Walter_2007_Open_Access_Management_Report.pdf

Jappe, A. (2007). Explaining international collaboration in global environmental change research. Scientometrics, 71(3), 367-390.

Jenkins, C.; Probets, S.G.; Oppenheim, C. and 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.

Kaltenborn, K. F. and Kuhn, K. (2004). The journal impact factor as a parameter for the evaluation of researchers and research. Revista Espanola de Enfermedades Digestivas, 96(7), 460-476.

Katz, J. S. and Martin, B. R. (1997), What is research collaboration? Research Policy, 26(1), 1-18.

Kim, M. J. (2005). Korean science and international collaboration, 1995-2000. Scientometrics, 63(2), 321-339.

Kim, K.-W. (2006). Measuring international research collaboration of peripheral countries: taking the context into consideration. Scientometrics, 66(2), 231-240.

Kirshop, B. and Chan, L. (2005). Transforming access to research literature for developing countries. Serials Review, 31(4), 246-255.

Klamer, A. and van Dalen, H. P. (2002). Attention and the art of scientific publishing. Journal of Economic Methodology, 9(3), 289-315.

Kling, R. and McKim, G. (2000). Not Just a Matter of Time: Field Differences in the Shaping of Electronic Media in Supporting Scientific Communication. Journal of the American Society for Information Science, 51(14), 1306-1320.

Lattin, J., Carroll, J.D., and Green, P.E. (2003). Analyzing multivariate data.

Laudel, G. (2002). Collaboration and reward. What do we measure by co-authorships. Research Evaluation, 11(1), 3-15.

Levine, M. (1965). Biology – one discipline or many? BioScience, 15(5), 346-348.

Lewison, G.; Anderson, J.; and Jack, J. (1995). Assessing track records. Nature, 377, 671.

Lor, P. J. and Britz, J. (2005). Knowledge production from an African perspective: International information flows and intellectual property. The International Information & Library Review, 37(2), 61-76.

Lorigo, L. and Pellacini, F. (2007). Frequency and Structure of Long Distance Collaborations in a Physics Community. Journal of the American Society for Information Science and Technology 58(10), 1497-1502.

Maffuli, N. (1995). More on citation analysis. Nature, 378, 760

Moed, H.F. (2005). Citation Analysis in Research Evaluation. Dordrecht, NL: Springer.

Nemeth, C. J. and Goncalo, J. A. (2005), Creative Collaborations From Afar: The Benefits of Independent Authors. Creativity Research Journal, 17(1), 1-8.

Nicholas, D.; Huntington, P. and Jamali, H. R. (2007). Open access in context: a user study. Journal of Documentation, 63(6), 853-878.

Nicholas, D.; Huntingdon, P. and Rowlands, I. (2005). Open access publishing: the views of some of the world's senior authors. Journal of Documentation, 61(4), 497-519.

O'Neill, J. (2000). The significance of an impact factor: implications for the publishing community. Learned Publishing, 13(2), 105-109.

Papin-Ramcharan, J. and Dawe, R. A. (2006). The Other Side of the Coin for Open Access Publishing - A Developing Country View. Libri, 56(1), 16-27.

Pelizzari E. (2003). Academic staff use, perception and expectations about open-access archives. A survey of social science sector at Brescia University. Available at: http://eprints.rclis.org/archive/00000737/01/Academic_staff_perception_about_Open_archives.htm

Ponds, R.; van Oort, F. G. and Frenken, K. (2007). The geographical and institutional proximity of research collaboration. Papers in Regional Science 86(3), 423-444.

Pouris, A. and Richter, L. (2000). Investigation into state-funded research journals in South Africa. South African Journal of Science, 96, 98–104.

Price, D. J. de Solla (1965). Networks of Scientific Papers. Science, 149(3683), 510-515.

Puliselic, L. and Petrak, J. (2006) Is it enough to change the language? A case study of Croatian biomedical journals. Learned Publishing, 19(4), 299-306.

Richardson M, Saxby C. Experimenting with open access publishing. Nature 2004. Available at: www.nature.com/nature/focus/accessdebate/12.html.

Rowlands, I.; Nicholas, D. and Huntingdon, P. (2004a). Journal publishing: what do authors want? Available at: http://www.nature.com/nature/focus/accessdebate/31.html

Rowlands, I.; Nicholas, D. and Huntingdon, P. (2004b). Scholarly communication in the digital environment: what do authors want? Findings of an international survey of author opinion: project report. London: Centre for Information Behaviour and Evaluation of Research, Department of Information Science, City University, 2004.

Royle, J.; Coles, L.; Williams, D.; and Evans, P. (2007). Publishing in international journals: An examination of trends in chinese co-authorship. Scientometrics, 71(1), 59-86.

Saiz-Sallinas, J. I. (1996). Failed professor. Nature, 381,186.

Schroter, S. and Tite, L. (2006). Open Access publishing and author-pays business models: a survey of authors' knowledge and perceptions. Journal of the Royal Society of Medicine, 99, 141-148.

Schroter, S.; Tite, L. and Smith, R. Perceptions of open access publishing: interviews with journal authors. BMJ, Published 26 January 2005

Shrum, W. and Campion, P. (2000). Are Scientists in Developing Countries Isolated? Science, Technology, and Society 5(1), 1-34.

Shrum, W. (1997). View from afar: 'Visible' productivity of scientists in the developing world. Scientometrics, 40(2), 215-235.

Shrum, W. (2005). Reagency of the Internet, or, how I became a guest for science. Social Studies of Science, 35(5), 723-754.

Schubert, A. and Glänzel, W. (2006). Cross-national preference in co-authorship, references and citations. Scientometrics, 69(2), 409-428.

Sotudeh, H. and Horri, A. (2007). Tracking open access journals evolution: Some considerations in open access data collection validation. Journal of the American Society for Information Science and Technology, 58(13), 2145-2156.

Swan, A. and Brown, S. (2005) Open access self-archiving: An author study. Technical Report, External Collaborators, Key Perspectives Inc.

Tijssen, R. J. W.; Mouton, J.; Van Leeuwen, T. N. and Boshoff, N. (2006). How relevant are local scholarly journals in global science? A case study of South Africa. Research Evaluation, 15(3), 163-174.

Tijssen, R. J. W. and Van Leeuwen, T. N. (2006). Measuring impacts of academic science on industrial research: a citation-based approach. Scientometrics, 66(1), 55-69.

Törmä, S., and Vakkari, P. (2004). Discipline, availability of electronic resources and the use of Finnish National Electronic Library - FinELib. Information Research , 10(1),

Paper No. 204. Available at: http://informationr.net/ir/10-1/paper204.html.

Tsay, M.-Y. (2006). Journal self-citation study for semiconductor literature: Synchronous and diachronous approach. Information Processing & Management, 42(6), 1567-1577.

van Dalen, H. P. and Henkens, K. (2005). Signals in science - On the importance of signaling in gaining attention in science . Scientometrics, 64(2), 209-233.

Velho, L. (1986), The meaning of citation in the context of a scientifically peripheral country, Scientometrics, 9(1-2), 71–89.

Vinkler, P. (1986). Evaluations of some methods for relative assessment of scientific publications. Scientometrics, 10, 157-178.

Wagner, C. S. (2005) Six case studies of international collaboration in science. Scientometrics 62(1), 3–26.

Wagner, C. S. and Leydesdorff, L. (2005). Network structure, self-organization, and the growth of international collaboration in science. Research Policy, 34(10), 1608-1618.

Wellcome Trust (2004). Costs and business models in scientific research publishing: a report commissioned by the Wellcome Trust. April 2004. Available at: www.wellcome.ac.uk/publications.

White, H. (2001). Authors as citers over time. Journal of the American Society for Information Science and Technology, 52(2), 87-108.

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

Xuemao, W. and Chang, S. (2006). Open Access – Philosophy, Policy and Practice: A Comparative Study. Proceedings of the World Library and Information Congress: 72nd IFLA General Conference and Council, Seoul, Korea. Available at: http://www.ifla.org/IV/ifla72/papers/157-Wang_Su-en.pdf.

Appendix 1. List of included journals in the study

Acta Biochimica Polonica
Acta Bioquimica Clinica Latinoamericana
Acta Protozoologica
Acta Zoologica.
Advances in Applied Microbiology
Advances in Biochemical Engineering - Biotechnology
African Entomology
African Zoology
Agricultural and Forest Entomology
American Journal of Hematology
American Journal of Primatology
American Museum Novitates
Animal Biology
Animal Biotechnology
Animal Genetics
Annales Zoologici
Annals of Applied Biology
Annals of Human Biology
Annual Review of Entomology
Annual Review of Genetics
Annual Review of Microbiology
Apidologie
Applied Biochemistry and Microbiology
Applied Entomology and Zoology
Archives of Biochemistry and Biophysics
Archives of Insect Biochemistry and Physiology
Archives of Microbiology
Behavior Genetics
Biochemistry
BioControl
Biological Chemistry
Biological Research
Biology Bulletin
Bioprocess and Biosystems Engineering
Bioscience
Bioscience, Biotechnology, and Biochemistry
Biotechnology Advances
Biotechnology Letters
BMC Biotechnology
BMC Evolutionary Biology
BMC Genetics
BMC Genomics
BMC Microbiology
Brazilian Archives of Biology and Technology
Brazilian Journal of Microbiology
Canadian Entomologist
Canadian Journal of Microbiology
Caribbean Journal of Science
Clinical Biochemistry
Clinical Microbiology and Infection
Contributions to Zoology
Critical Reviews in Biotechnology
Current biology
Current Microbiology
Current Opinion in Chemical Biology
Current Opinion in Genetics & Development
Developmental Biology
Electronic Journal of Biotechnology

Entomologica Fennica
Entomological News
Ethology
European Journal of Entomology
Experimental & Molecular Medicine
FASEB Journal
Florida Entomologist
Folia Parasitologica
Folia Zoologica
Food Biotechnology
Food Microbiology
Food Technology and Biotechnology
Frontiers in Bioscience
Genes & Genetic Systems
Genes, Brain and Behavior
Genetics and Molecular Biology
Genome
Genomics
IEEE Transactions on Information Technology in Biomedicine
Indian Journal of Biochemistry and Biophysics
Insect Molecular Biology
Insectes Sociaux
International Journal of Acarology
International Journal of Molecular Sciences
International Microbiology
Italian Journal of Zoology
Journal of Animal Ecology
Journal of Applied Entomology
Journal of Basic Microbiology
Journal of Biochemistry and Molecular Biology
Journal of Biomedicine and Biotechnology
Journal of Biosciences
Journal of Chemical Technology and Biotechnology
Journal of Clinical Microbiology
Journal of Genetics
Journal of Insect Behavior
Journal of Insect Physiology
Journal of Lipid Research
Journal of Microbiology
Journal of Microbiology and Biotechnology
Journal of Molecular Catalysis B: Enzymatic
Journal of Pest Science
Journal of Plant Biochemistry and Biotechnology
Journal of Proteome Research
Korean Journal of Genetics
Laboratory Animals
Medical & Veterinary Entomology
Microbiology and Immunology
Microbiology and Molecular Biology Reviews
Molecular and Cellular Biology
Molecular Biology
Molecular BioSystems
Molecular Vision
Mutagenesis
Nature Biotechnology
Neurogenetics
North American Journal of Aquaculture
Nucleic Acids Research
Physiological Entomology

PLOS Biology
Process Biochemistry
Raffles Bulletin of Zoology
Russian Journal of Genetics
Steroids
Structure
Systematic Biology
Trends in Biochemical Sciences
Trends in Biotechnology
Trends in Microbiology
Zoo Biology
Zoologicheskii Zhurnal
Zoosystema

Appendix 2. Inter-correlations of the independent variables in the first model.

	2	3	4	5	6	7	8	9	10	11
1 JIF	-.13	-16*	-.22*	.44	.10	.06	.02	-.32*	.49*	-.11
2 OA		-.17*	.02	-.01	.004	.04	.02	.38*	-.13	.05
3 Entomology			-.14	-19*	-.21*	-.19*	-.19*	-.11	-.09	-.07
4 Zoology				-.19*	-.26*	-.13	-.19*	-.02	-.17*	.11
5 Microbiology					-.01	-.01	-.13	.05	.18*	-.07
6 Biochemistry						-.06	.07	.06	.02	.11
7 Genetics							-.13	-.004	.06	-.07
8 Biotechnology								.03	.11	-.07
9 Geographic location of journal: developing country									-.17	.26
10 Share of reviews										-.03
11 Share of articles not in English										

* $p < .05$ level (two-tailed).

Appendix 3. Inter-correlations of the independent variables of the second model.

* p < .05 (two-tailed).

	2	3	4	5	6	7	8	9	10	11	12	13
1 Author(s): developing countries	-.21*	-.13*	.00	-.12	-.07	.01	.15*	.01	.11	.17*	-.05	-.05
2 Author(s): developed and developing countries		.09	.01	-.08	-.07	-.00	.04	-.06	.03	-.12*	-.04	-.06
3 JIF			-.11	-.20*	-.16*	-.03	.13*	.03	-.03	-.27*	.45*	-.10
4 OA				.01	-.15*	-.05	-.00	.11	.00	.24*	-.13*	.04
5 Entomology					-.10	-.18*	-.25*	-.08	-.18*	.02	-.15*	.10
6 Zoology						-.17*	-.21*	-.16*	-.17*	-.13*	-.08	-.05
7 Microbiology							-.06	-.00	-.10	.11	.05	-.06
8 Biochemistry								-.09	-.01	.02	-.03	.07
9 Genetics									-.16*	.03	-.02	-.06
10 Biotechnology										-.04	.18*	-.06
11 Geographic location of journal: developing country											-.14*	.27
12 Share of reviews												-.03
13 Share of articles not in English												

11. The effects of open access on un-published documents:

A case study of economics working papers

Tove Faber Frandsen
Royal School of Library and Information Science
Birketinget 6, DK-2300 Copenhagen S, DENMARK
Phone: +45 3258 6066
Fax: +45 3284 0201
E-mail: tff@db.dk

Abstract

The use of scholarly publications that have not been formally published in e.g. journals is widespread in some fields. In the past they have been disseminated through various channels of informal communication. However, the Internet has enabled dissemination of these unpublished and often unrefereed publications to a much wider audience. This is particularly interesting seen in relation to the highly disputed open access advantage as the potential advantage for low visibility publications has not been given much attention in the literature. The present study examines the role of working papers in economics during a ten-year period (1996 to 2005). It shows that working papers are increasingly becoming visible in the field specific databases. The impact of working papers is relatively low; however, high impact working paper series have citation rate levels similar to the low impact journals in the field. There is no tendency to an increase in impact during the ten years which is the case for the high impact journals. Consequently, the result of this study does not provide evidence of an open access advantage for working papers in economics.

Introduction

Open access (OA) based data sources provide new opportunities for performing citation analysis. They are interesting as an alternative to the traditional citation databases primarily due to coverage. OA based data sources are not restricted to journal articles as the traditional citation databases implying that some disciplines characterised by many e.g. working papers and monographs may be better analysed using a data source based on OA resources. However, 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 even if they do 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 perceived major benefit for authors making their publications available OA is the higher number of citations to OA publications presumably due to higher visibility and /or access. Lawrence (2001) analysed the effects of OA on citation impact for a sample of conference documents and found freely available papers to have greater impact. However, his study was restricted only to computer science which made it difficult to generalise. This effect, "the open access advantage" has since been confirmed on larger samples by Antelman (2004), Harnad and Brody (2004a), Hajjem et al. (2005), and Hajjem, Harnad and Gingras (2005), Metcalfe (2005), Henneken et al. (2006). However, one should be careful making causal arguments as pointed out by e.g. Craig et al. (2007) stressing that in order to conclude that OA publication causes more citations, we need data to illustrate causation, not just association. The increased number of citations could be caused by other factors than the increased visibility of the work as stressed in the work by Kurtz et al. (2005), Davis and Formerth (2006), Moed (2007), Davis (2009), Davis et al. (2008) and Norris, Oppenheim and Fytton (2008). 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 non-OA, show no significant difference in citation rates when he controls for the effects of these factors. Davis and Fromerth (2006) find that the OA advantage is not a result of more citations to all the OA papers

but rather of self-selection (authors select their best papers to self-archive). Gaule and Maystre (2008) find a very small open access effect which to a great extent is due to a self-selection effect rather than a diffusion effect.

The existing analyses of the open access advantage have mainly focused on comparing citation data for archived and not archived journal articles (e.g. Metcalfe, 2006; Moed, 2008; Norris, Oppenheim & Rowland, 2008) or OA and non-OA articles in hybrid journals (Eysenbach, 2006; Davis, 2009). Little attention has been paid to comparing citation data for publications in earlier stages than the journal article. Some of the studies include earlier versions of the publication when determining the OA status of a publication as they do not distinguish between various versions of publications (e.g. Antelmann, 2004; Norris, Oppenheim & Rowland, 2008). Others include citations to these publications as means to ensure a fixed time window for the citations (e.g. Moed, 2007). Although earlier versions of the journal publications can be included in a pool with the journal articles the focus is on the journal article version of publications. Few related studies exist. As already mentioned, Lawrence (2001) found evidence of an open access advantage of computer science conference proceedings. However, in computer science conference papers are considered the final product (Kling & McKim, 1999; Goodrum et al., 2001) and are often even more prestigious than journals. Schwarz and Kennicutt (2004) and Metcalfe (2006) did studies in the field of astronomy finding evidence of an open access advantage for conference proceedings, although the citation rates are very low regardless of being available open access or not. However, results by Kurtz and Henneken (2007) show no evidence of an open access advantage within the same field for journal articles which is explained by Harnad (2006) as a result of astronomy being a special case because "all active, publishing researchers already have online access to all relevant journal articles".

The open access advantage of publications in earlier stages than the formal publication is particularly interesting as such publications have had limited visibility before the Internet enabled wider dissemination. The Internet has made it possible to disseminate the earlier versions of publications worldwide and research is made available on e.g. personal and institutional websites. Consequently, there has been a growth in both the number (Meadows, 1998: 164) and forms (Farace, 1997). The aim of this study is to investigate if un-published economics literature increases in numbers, visibility and consequently citations over a ten-year period. The study can contribute to the ongoing

discussion of open access advantage as this study focuses on a different publication type than the existing studies and includes the perspective of the development of a potential open access advantage over time (the developments over time have recently been included in a study by Davis, 2009). Furthermore, the study contributes to the discussion of weighting schemes of publications. An increased amount of grey literature and the easy access may result in a collapse of the distinction between published and un-published literature (Banks, 2006a) which may be cause for concern in terms of the measures typically used to analyse research (e.g. citation rates). In addition, such a collapse would result in increasing recognition and thus citations rates. Should the distinction be collapsing equal credit should be assigned to un-published and published publications.

The paper is structured as follows: the next section describes and discusses the chosen methods followed by the results. Finally, the paper is completed with a discussion and conclusion.

Methods

Economics was chosen for the case study because the working paper (WP) is a central document type within this field. The tradition of paper manuscript publication in economics goes back to the 1960s (Kling, Spector & McKim, 2002) and their importance within the field is well established within the literature (Whitley, 1991; Pierce, 1992; Robinson & Poston, 2004; Zhang, 2007).

The present study includes two separate analyses. First, the share of grey literature of the publications was determined by calculating the percentage of publications in field specific databases made up by working papers. Two major economics databases were used: EconLit, which includes primarily books, WPs and journal articles and RePEc, which is an open access based resource. In EconLit the analyses could be done delineating to specific publication years and thus provide data on the development during the last two or three decades. In RePEC it was possible to delineate to publications updated within the last year (and the data collection took place January 2nd 2008) and consequently, this data must be compared with the data on 2007 from

EconLit. This analysis provides evidence of visibility and does consequently not provide evidence for the actual numbers and shares of WPs.

Secondly, a sample of economics WP series from 1996 to 2005 (about 2000 WPs in total) was analysed to detect a possible increase in citations to WPs (the WP series were made available open access continuously over a period from 1998 to 2000). A list of the included WPs is available in table 1. The citation window was set to three years and consequently for a WP from 1996 citations from 1996 to 1998 was included. The WPs were looked up individually in the citation indexes as there is no consistent assignment of cited works for WPs.

Working papers	Publications included in the study
Boston College Working Papers in Economics	305
Cambridge Working Papers in Economics	372
Centre for Economic Policy Research, Research School of Social Sciences, Australian National University	169
Federal Reserve Bank of Minneapolis. Working papers	80
Harvard Institute of Economic Research Working Papers	35
School of Finance and Economics, University of Technology, Sydney	88
Universitat de Barcelona. Espai de Recerca en Economia. Working Papers in Economics	145
University of California at Los Angeles (UCLA). Economics Working papers	99
University of Copenhagen. Department of Economics. Discussion papers.	251
University of Rochester. RCER Working Papers.	103

Table 1. Overview of included working papers

Impact factors (IFs) tend to increase in general over time and thus a reference sample was needed. The reference sample consisted of 30 randomly selected journals in the economics subject category in JCR®. Some journals may be included or excluded from the subject category (alternatively moved to another subject category) during the investigated period and thus the sample is drawn among the journals present in the subject category throughout the entire period. The sample consists of about 1/3 of the population (some journals are only present in the beginning of their indexed period with an exceptionally limited number of publications and are thus not suited to be included). The journal impact factors (JIFs) for journals were computed correspondingly to the computation of WP IFs. Citations to all publications were included in the numerator, however, only the following publication types were included in the denominator: article, review, letter, note. Preferably, only citations to articles in numerator are included but due to the indexing policy of the citation indexes, citations to all publication types are included in the numerator. A list of the included journals is available in table 2.

Journals	Articles included in the study
Bulletin of Indonesian Economic Studies	163
Canadian Journal of Economics	670
Contemporary Economic Policy	421
Ecological Economics	1013
Econometrica	594
Economic Development and cultural Change	358
Economic Development Quarterly	270
Economic Geography	191
Economic Journal	779
Economica	342
Ekonomicky Casopis	580
Health Economics	491
Hitotsubashi Journal of Economics	654
International Economic Review	102
Jahrbücher für Nationalökonomie und Statistik	431
Journal of Development Economics	672
Journal of Econometrics	861
Journal of Economic Literature	179
Journal of Economic Theory	951
Journal of Economics & Management Strategy	240
Journal of Environmental Economics and Management	484
Journal of Evolutionary Economics	230
Journal of Industrial Economics	242
Journal of International Economics	550
Journal of Labor Economics	337
Journal of Real Estate Finance and Economics	358
Oxford Bulletin of Economics and Statistics	375
Review of Income and Wealth	280
Scandinavian Journal of Economics	346
South African Journal of Economics	372

Table 2. Overview of included journals

The development over time in IF of journals and WP series is analysed graphically. The journals as well as the WP series are separated in two according to impact factor. This is due to the fundamentally very different development over time by high impact and low impact journals which will be evident in the results section below.

Results

Figure 1 shows an increase in the relative size of WPs. Their percentage of the publications in EconLit has increased from about 7 percent in 1980 to about 16 percent in 2006. This study analyses data in the ten-year period from 1996 to 2005 and in this period the share of working papers is increasing from about 10 percent to 16 percent.

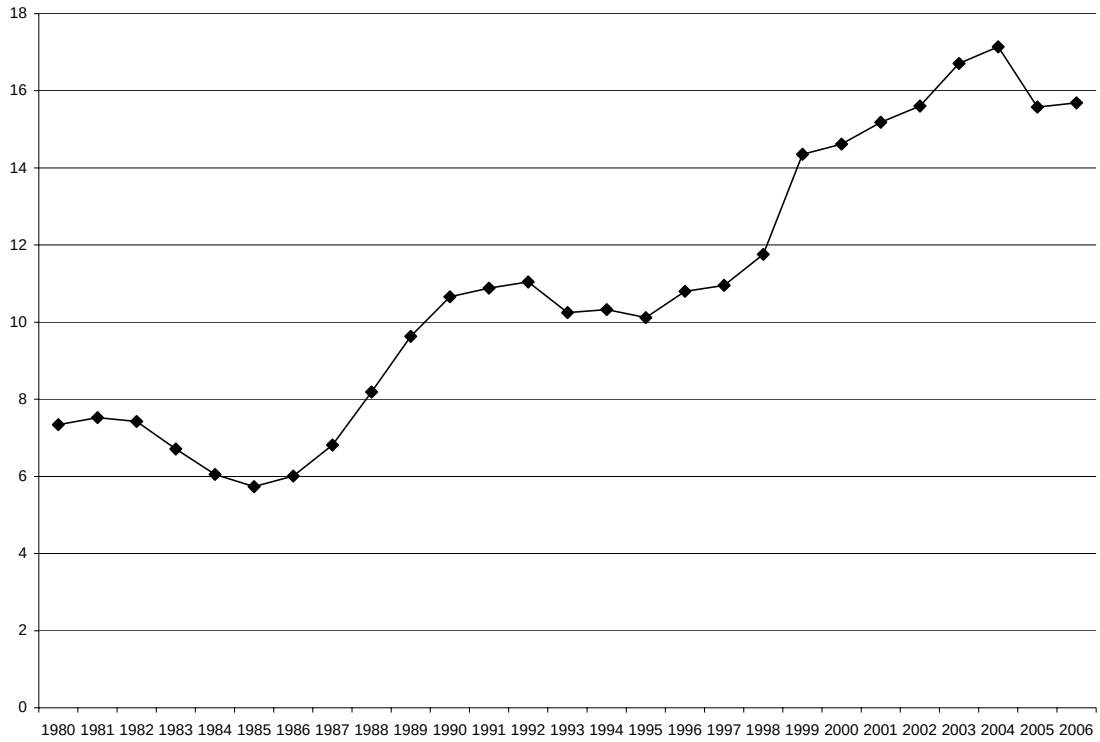


Figure 1. WPs as a share of the total publications in EconLit. Percentages are shown as moving averages of three years.

Consequently, working papers make up an increasing share of the publications indexed in EconLit, although the figures in figure 1 are made up by a distribution of WPs that differs among different subject areas. It is evident in table 3 that the importance of WPs varies substantially. For some subject areas the percentage of WPs is as low as 3 percent whereas in others it is as high as 38 percent.

Subject areas	**EconLit**	**RePEc**
General Economics and Teaching	11	72
Schools of Economic Thought and Methodology	3	67
Mathematical and Quantitative Methods	38	76
Microeconomics	17	87
Macroeconomics and Monetary Economics	17	85
International Economics	13	90
Financial Economics	10	91
Public Economics	17	87
Health, Education, and Welfare	13	87
Labor and Demographic Economics	15	81
Law and Economics	11	92
Industrial Organization	8	84
Business Administration and Business Economics; Marketing; Accounting	5	83
Economic History	14	97
Economic Development, Technological Change, and Growth	7	83
Economic Systems	3	86
Agricultural and Natural Resource Economics; Environmental and Ecological Economics	7	85
Urban, Rural, and Regional Economics	6	89

Table 3. Percentages of WPs in EconLit (2007) and RePEc (2007).

Furthermore, table 3 shows that the percentages of WPs in RePEc and thus the importance are much higher in RePEC. The WPs make up from 67 to 97 percent of the publications indexed in the database. Consequently, the share of WPs in RePEC is from 2 to more than 25 times higher than in EconLit and thus WPs are much more visible in

the open access based resource. It should be noted that some of the subject areas with the highest percentages of WPs in EconLit could be one of the lowest in RePEc. An example of such subject area is Law & Economics with 11 percent WPs in EconLit and 92 in RePEc. Mathematical and Quantative Methods have the highest share in EconLit (38) but one of the lowest in RePEc (76). These relative differences are due to the different indexing policies of EconLit and RePEc. EconLit includes selected WP series listed on RePEc. RePEc includes WP series as well as WPs not necessarily in WP series. An author can self-archive several versions of the same WP and in some cases there as many as 5 or more versions of the same WP. Although there are relative differences between the two databases it is clear that RePEc is dominated by WPs. The important question is then if this visibility results in an increased attention and thus citations.

Figure 2 depicts the development of average impact factors of 10 WP series. For reasons of comparison the figure also includes the IFs of 30 economics journals. IFs are calculated using diachronous IFs with a 1-year publication period and a 3-year citation period. The JIFs are shown excluding journal self-citations to ensure a reasonable comparison. The WPs are not included as source documents in the citation indexes and consequently the citation rates of WPs are calculated excluding self-citations. There is a clear tendency of the journals to increase their average impact factor through this ten year period and the WP series have little or no increase. However, this comparison is not quite reasonable as there are huge differences between low impact and high impact journals and WP series.

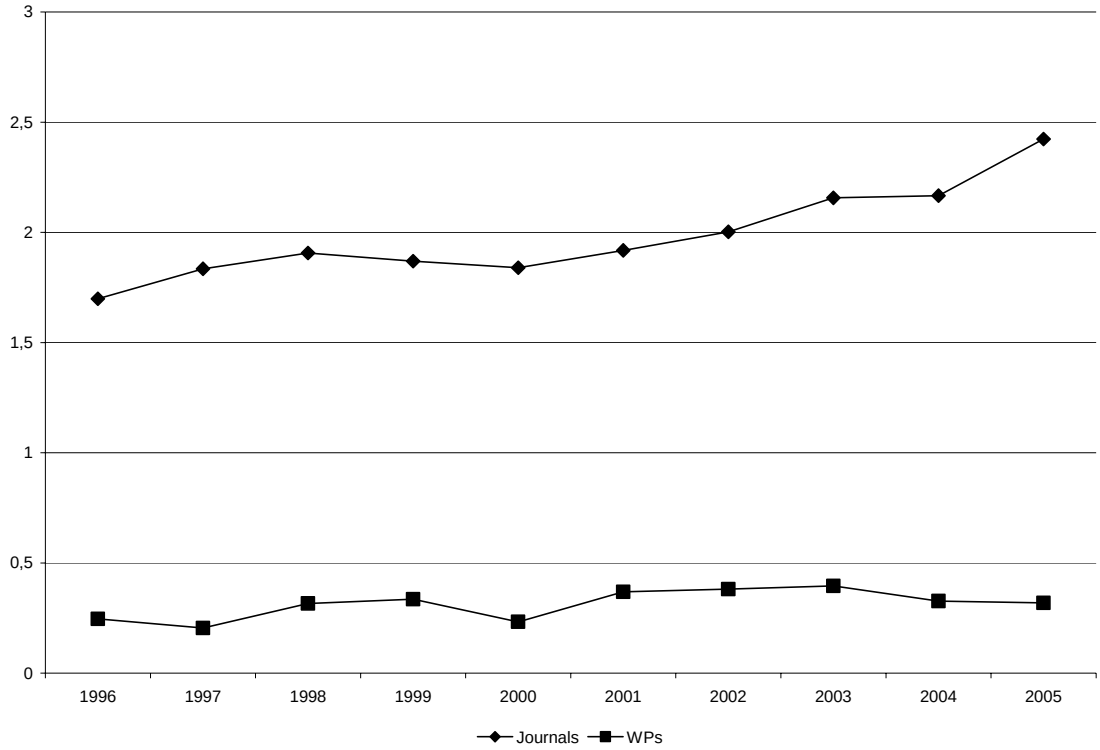


Figure 2. Impact factors of WP series and journals.

Figure 3 depicts the development of average impact factors of WP series separated in two groups: high impact WP series (1210 WPs in total) and low impact WP series (752 WPs in total). For reasons of comparison the figure also includes the IFs of economics journals also separated in two groups: high JIF journals (5298 publications in total) and low JIF journals (8238 publications in total). The division is made on the basis of the average IF of both WP series and journals. The set of journals is split in two at an average JIF of about 1.5 and the WP series are split at an average IF of about .3.

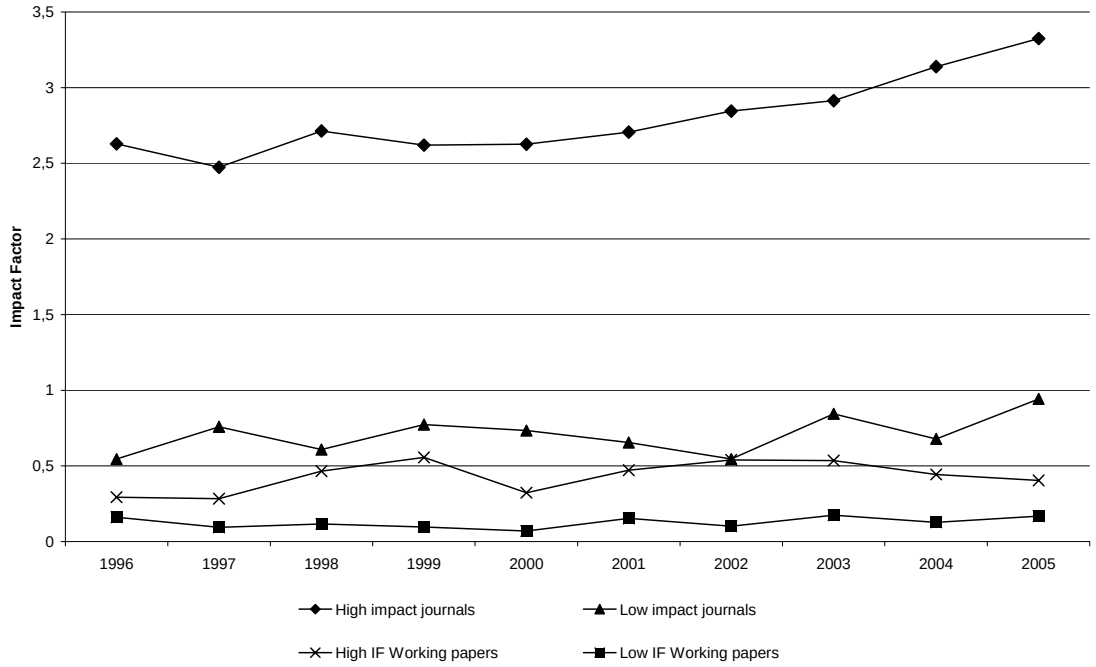


Figure 3. Impact factors of WP series and journals divided into low and high impact.

The key issue in this figure is not the difference in level of IFs but rather the development over time. It is evident that the high impact journals increase their IFs considerably. The IFs of the low impact journals do not experience the same increase, if any increase at all. The citation numbers of the low impact journals are smaller causing greater fluctuations which make it more difficult to determine the magnitude of a possible increase in JIF. However, figure 3 depicts the development in impact of high impact working paper series to be equivalent at best to that of low impact journals. Furthermore, the impact of low impact WP series is stable over time with a diminutive increase at best. During this 10 year period all of the included WP series were made available OA (some over the period of a few years), however, the increased visibility did not result in a dramatic increase in citation rates. The possible slight increase is not impressive, especially in comparison to that of the traditional journals.

Discussion and Conclusion

The results of this study relate to the literature on credit assigning and open access advantage. Consequently, these two themes form the structure of this discussion.

Bibliometric assessment of research performance is based on the research output of those being assessed, regardless of the exercise being based on publication or citation analysis. Research output forms the basis, and thus an operational definition of the term is required as it can be defined in numerous ways. Research output can be defined relatively restricted as the number of research articles, notes, letters and reviews published in journals covered by the citation indexes (e.g. Van Raan, 2004). A broader definition of research output is all peer-reviewed journal articles (e.g. Neri & Rodgers, 2006; Rodgers & Neri, 2007) and an even more including definition is journal articles, contributions to books and monographs (e.g. Van der Meulen & Leydesdorff, 1991). However, research output can be defined to entail other channels of publishing. Scholars can disseminate their research via non-scholarly publications directed at the general public (termed the “fourth literature of social science” by Hicks (1999, 2004)). Another publishing channel is the so-called grey or gray literature.

Grey literature is generally not peer reviewed but can be and un-refereed publications are not necessarily grey. The term grey literature is used as means to describe a rather heterogenic group of documents sharing some characteristics determined by the definition of the term. Grey literature is included in the UNISIST model of scientific and technical communication as formal, unpublished communication (UNISIST, 1971) and furthermore the revision of the model from 2003 includes preprint databases as distributors of grey literature (Søndergaard, Andersen, & Hjørland, 2003). Numerous definitions of grey literature exist, although, the definitions by the International Conferences on Grey Literature are often used (e.g. Benzies et al., 2006; Søndergaard, Andersen, & Hjørland, 2003). The Sixth International Conference on Grey Literature defines grey literature as "Information produced on all levels of government, academics, business and industry in electronic and print formats not controlled by commercial publishing i.e. where publishing is not the primary activity of the producing body." (www.greynet.org, accessed April 2008). Some definitions of grey literature are tied to the lack of accessibility of the publications (Auger, 1998). Grey scientific and technical literature includes conference literature, technical reports, theses, dissertations, patents

and government publications (Walker & Hurt, 1990). Grey literature also includes working papers (Auger, 1998) and within the last few decades electronic preprint archives have become an important dissemination channel for grey literature (Luzi, 2000) although distribution of preprints is well-known in studies of dissemination of research going back to Garvey and Griffith (1963).

Bibliometrics for evaluation purposes typically distinguishes between refereed and un-refereed documents and/or published and un-published documents. Un-refereed and unpublished documents would rarely receive the same recognition. Grey literature is generally not peer-reviewed (Banks, 2006b) which causes Archambault and Vignola-Gagné (2004) to note: "caution must be the watchword here, because this type of literature should be considered at least in part as an SSH [social sciences and humanities] dissemination medium outside the academic community instead of a scientific communication medium". Moreover, Meadows (1998: 164) notes that the "refereed journal articles and scholarly monographs are still regarded as the definitive statements of the results of research projects". Consequently, grey literature and un-refereed documents are typically excluded from assessments of research performance. However, in assessment of research performance research output has been defined including both other scholarly and non-scholarly publications. As means to perform bibliometric analyses of the humanities Moed, Luwel and Nederhof (2002) includes all publications reported by the members of faculty and then divide into substantial research contributions, small contributions, publications for the general public and other types of publications. Nederhof and van Raan (1993) divides into articles, books, book chapters and other types of publication. Ho (1998) includes all publications and applies a weighting scheme in which forms of grey literature are assigned weights of 2 per cent or less of an international journal article.

The little recognition of grey literature is apparent in citation analyses as grey literature is typically not heavily cited (Alberani, Pietrangeli, and Mazza, 1990; Nederhof & Noyons, 1992; Pelzer & Wiese, 2003; Salman et al., 2007); although there are field differences in the use of grey literature (Pelzer & Wiese, 2003).

The results of this study also reveal low citation rates of the WPs and there is no clear tendency to an increase over time. However, there are no signs of a decrease over time either as found by Lisée, Larivière and Archambault (2008) in the case of conference

proceedings. Furthermore, the results of this study indicate that there is a relatively modest difference in citation rates of low impact journals and high impact WP series. The difference between the two in citation rates seems to be relatively stable throughout the ten-year period, although with some fluctuations. The clear-cut boundary between published and unpublished publications as well as refereed and un-refereed publications seems less clear-cut as measured by citation rates. Kling and McKim (2000) and Kling (2004) stress that scholarly publishing is a continuum in the paper-only world as well as in the electronic and scholarly communication varies considerably across fields. Field specific weighting schemes could be worth considering as means to capture differences across fields in the recognition of specific publication types. Economics appears to be a field with relatively high recognition of working paper even though it is typically un-refereed and a form of grey literature.

The second theme of this discussion, the open access advantage, needs an initial clarification of the relations to other measures of usage. The supposed increase in the number of citations received by open access publications compared to (similar) non-OA publications must be kept separated from other measures of usage. Open access would, all other things equal, have a greater potential audience. It would at least be difficult to imagine a smaller audience for a specific journal article if self-archived in an open access repository. An example of other measures of usage is number of downloads (e.g. Bollen and Van de Sompel, 2008; Bollen et al., 2006). The number of citations and the number of downloads are not necessarily causally related meaning that increased potential audience and thus readership does not necessarily imply a greater number of citations. However, there may relations between the measures in general (e.g. Perneger, 2004; Brody, Harnad & Carr, 2006). Davis et al. (2008) isolates the effects of OA for OA-articles and non-OA articles and finds that OA-articles are more downloaded than non-OA articles in the same journals. They are, however, not more cited (although a study with a longer citation window would have been preferred).

A confirmation of the open access advantage (especially if the size of the effect would be found to be large) would have implications on the fundamental theories of scholarly communication and bibliometrics. It would imply that scholars cite specific publications simply because they are easily available which lend support to theories of citing that emphasize the behaviour of scholars as a balance between cost and benefit and consequently, question both the normative theory of citing and other theories that are

based on the principle of least effort (the reader is referred to Nicolaisen, 2007 for a recent review of theories of citing).

The great challenge of studies of open access is the determination of causation. A study that illustrates association is much simpler to perform but do not prove or disprove the existence of the open access advantage. In order to determine causation a study must control for the effect of other variables such as early view and quality bias. Furthermore, determining the status of a publication as being OA or non-OA is not a trivial task. A publication can be made OA in numerous ways and the publication can exist in an earlier version maybe even with a different title. Finally, there are issues of field specific variations which complicate studies across fields. Adding all these time consuming challenges it is tedious work to determine causation of a potentially very small effect. Furthermore, a considerable data material is necessary to be able to control for the many variables (see e.g. Norris, Oppenheim and Rowland (2008) as an example of a study including more than 4,600 articles and yet not having enough to determine the cause).

This study does not claim to determine causation but rather point to potential developments over time in the citation rates of open access publications that have not previously been focus of attention. Firstly, the development over time does not indicate that the WP series have received more citations due to the transformation from print to online version during the years 1998 to 2000. Should there be a positive tendency; the size of the effect is diminutive. Secondly, the relative constant citation rates of these WP series are noteworthy as this type of publication has been almost invisible and very hard to obtain before the Internet. One would imagine that the effect would be considerable for these publications as they have had limited visibility. This is, however, in accordance with the results found by Schwarz and Kennicutt (2004) on the open access advantage of conference proceedings. They stress that increased visibility is not necessarily a guarantee of increased citation rates.

> [P]reprint posting increases the relative visibility of non-peer-reviewed papers by a comparable factor, but the factor-of-20 difference between proceedings papers and ApJ [The Astrophysical Journal] papers remains the same regardless of whether the respective papers are posted on astro-ph [the arXive electronic preprint server] or not. This should serve as a caution to anyone who might

> believe that preprint posting alone is sufficient to assure that a paper is widely recognized and cited.

It is worth keeping in mind, however, that citations is not the only measure of usage. As already stated, the tradition and importance of working papers in economics is well documented. Nonetheless, when turning working papers into journal articles authors tend to prefer published documents over unpublished documents in their reference lists (Frandsen & Wouters, 2009). This indicates that although used and often cited authors may prefer to cite peer reviewed articles if possible which then again may imply that readership is to be kept separate from citation rates. Open access may increase readership but not citation rates as argued by e.g. Craig et al. (2007) and Davis et al. (2008). Although, the open access advantage defined in terms of citation rates is the present focus, it can be argued that increased readership is as important as citation rates (Bognolo, 2008; Latronico, 2008). Lucas and Willinsky (in press) even agitate for open access from a perspective of democracy. Although, analyses of the use of open access based resources in various stages of a research project are indeed interesting, it is beyond the scope of this study.

Summing up, the results of this study have shown that the impact of working papers is relatively low in the field of economics. It is worth noting, however, that high impact working papers have levels of citation rates similar to the low impact journals. There is no clear tendency to an increase in impact during the ten years, which is the clear tendency of the high impact journals. Consequently, the result of this study does not provide evidence of an open access advantage for working papers in economics. The results of this study do raise the question of field specific weighting schemes of publications as publications that typically receive little or no credit in bibliometric assessment of research performance may be highly recognised in some fields.

References

Alberani, V.; Pietrangeli, P.D.C. and Mazza, A.M.R. (1990). The use of grey literature in health sciences: a preliminary survey. Bulletin of the Medical Library Association, 78(4),358-363.

Antelman, K. (2004). Do open-access articles have a greater research impact? College and Research Libraries, 65, 372-382

Archambault, E., and Vignola-Gagné, E. (2004). The Use of Bibliometrics in the Social Sciences and Humanities. Prepared for the Social Sciences and Humanities Research Council of Canada (SSHRC). http://www.sciencemetrix.com/pdf/SM_2004_008_SSHRC_Bibliometrics_Social_Science.pdf.

Auger, C. P. (1998). Information sources in grey literature, 4th edition. London: Bowker Saur.

Banks, M. (2006a). Towards a continuum of scholarship: the eventual collapse of the distinction between grey and non-grey literature? In Farace, D. J., Frantzen, J. (Eds),GL7 Conference Proceedings, 7th International Conference on Grey Literature: Open Access to Grey Resources, Nancy, 5-6 December, TextRelease, Amsterdam.

Banks, M. (2006b). Connections between open access publishing and access to gray literature. J Med Libr Assoc, 92(2), 164-166.

Benzies, K., Premji, S., Hayden, K. A., and Serrett, K. (2006). State-of-the-evidence reviews: Advantages and challenges of including grey literature. Worldviews on Evidence-Based Nursing, 3, 55-61.

Bognolo, G. (2008). Access is more important than citation. Rapid response to Davis, P.; Lewenstein, B.; Simon, D.; Booth, J. & Connolly, M. (2008). Open access publishing, article downloads, and citations: randomised controlled trial. British Medical Journal 337:a568. http://www.bmj.com/cgi/eletters/337/jul31_1/a568#199806.

Bollen, J. and van de Sompel, H. (2008). Usage Impact Factor: The effects of sample characteristics on usage-based impact metrics. Journal of the American Society for Information Science and Technology, 59(1), 136-149.

Bollen, J.; Marko A. R. and Van de Sompel, H. (2006). Journal status. http://www.citebase.org/cgi-bin/fulltext?format=application/pdf&identifier=oai:arXiv.org:cs/0601030

Brody, T.; Harnad, S. and Carr, L. (2006). Earlier web usage statistics as predictors of later citation impact. Journal of the American Society for Information Science and Technology, 57(8), 1060-1072.

Craig, I.; Plume, A.; McVeigh, M.; Pringle, J. and 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. (2009). Author-choice open access publishing in the biological and medical literature: a citation analysis. Journal of the American Society for Information Science and Technology.60(1), 3–8.

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

Davis, P.; Lewenstein, B.; Simon, D.; Booth, J. & Connolly, M. (2008). Open access publishing, article downloads, and citations: randomised controlled trial. British Medical Journal 337:a568

Eysenbach, G. (2006) Citation Advantage of Open Access Articles PLoS Biology, 4(5).

Farace, D.J. (1997). Rise of the phoenix: A review of new forms and exploitations of grey literature. Publishing Research Quarterly, 13(2), 69-76.

Frandsen, T. F. & Wouters, P. (2009). Turning working papers into articles: An exercise in micro-bibliometrics. Journal of the American Society for Information Science and Technology, 60(2).

Garvey, W.D., & Griffith, B.C. (1963). An overview of the structure, objectives, and findings of the American Psychological Association's project on scientific information exchange in psychology. Washington, DC: American Psychological Association.

Gaule, P. and Maystre, N. (2008). Getting cited: does open access help? CEMI-WORKINGPAPER-2008-007. http://ilemt.epfl.ch/repec/pdf/cemi-workingpaper-2008-007.pdf

Goodrum, A.A., McCain, K.W., Lawrence, S. and Giles, C.L. (2001). Scholarly publishing in the Internet age: a citation analysis of computer science literature, Information Processing and Management, 37(6), 661–675.

Hajjem, C., Gingras, Y., Brody, T., Carr, L. and Harnad, S. (2005) Open Access to Research Increases Citation Impact. Technical Report, Institut des sciences cognitives, Université du Québec à Montréal.

Hajjem, C., Harnad, S. and Gingras, Y. (2005) Ten-Year Cross-Disciplinary Comparison of the Growth of Open Access and How it Increases Research Citation Impact. IEEE Data Engineering Bulletin 28(4), 39-47.

Harnad, S. (2006). The special case of astronomy. Open Access Archivangelism, October 14. 2006.

Harnad, S. and Brody, T. (2004). Comparing the impact of open access (OA) vs. non-OA articles in the same journals. D-Lib Magazine, 10.

Henneken, E. A., Kurtz, M. J., Eichhorn, G., Accomazzi, A., Grant, C., Thompson, D., and Murray, S. S. (2006). Effect of E-printing on Citation Rates in Astronomy and Physics Journal of Electronic Publishing, Vol. 9, No. 2, Summer 2006, also in ArXiv, Computer Science, cs.DL/0604061, v2, 5 June 2006 http://arxiv.org/abs/cs/0604061

Hicks D. (1999). The difficulty of achieving full coverage of international social science literature and the bibliometric consequences. Scientometrics. 44(2): 193-215.

Hicks D. (2004). The Four Literatures of Social Science. Handbook of Quantitative Science and Technology Research. Ed.: Henk Moed. Dordrecht: Kluwer Academic.

Ho K.K. (1998). Research output among the three faculties of business, education, humanities & social sciences in six Hong Kong universities. Higher Education. 36(2): 195-208.

Kling, R. and McKim, G. (1999). Scholarly communication and the continuum of electronic publishing. Journal of the American Society for Information Science, 50, 890-906.

Kling, R. and McKim, G. (2000). Not just a matter of time: Field differences and the shaping of electronic media in supporting scientific communication. Journal of the American Society for Information Science, 51, 1306-1320.

Kling, R. (2004). The internet and unrefereed scholarly publishing. Annual Review of Information Science and Technology, 38, 591-631.

Kling, R., Spector, L., & Mckim, G. (2002). Locally controlled scholarly publishing via the internet: The guild model. The Journal of Electronic Publishing, 8(1).

Kurtz, M.J.; Eichhorn, G.; Accomazzi, A.; Grant, C.; Demleitner, M.; Henneken, E. and Murray, S.S. (2005). The effect of use and access on citations. Information Processing and Management: an International Journal, 41(6), 1395-1402.

Kurtz, M.J. and Henneken, E. (2007). Open Access does not increase citations for research articles from The Astrophysics Journal. arXiv:0709.0896v1

Latronico, N. (2008). Who cares for citation advantage? Rapid response to Davis, P.; Lewenstein, B.; Simon, D.; Booth, J. & Connolly, M. (2008). Open access publishing, article downloads, and citations: randomised controlled trial. British Medical Journal 337:a568. http://www.bmj.com/cgi/eletters/337/jul31_1/a568#199806.

Lawrence, S. (2001) Online or Invisible? Nature 411 (6837): 521. http://www.neci.nec.com/~lawrence/papers/online-nature01/.

Lisée, C.; Larivière, V. and Archambault, E. (2008). Conference proceedings as a source of scientific information: A bibliometric analysis. Journal of the American Society for Information Science and Technology, 59(11):1776–1784.

Lucas, R. and Willinsky, J. (In Press). Open Access to E-Research. In N. Jankowski (Ed.), e-Research: Transformations in Scholarly Practice. London: Routledge.

Luzi, D. (1998). E-print archives: A new communication pattern for grey literature. Interlending and Document Supply 26(3), 130-139.

Meadows, A. J. (1998). Communicating Research. San Diego, CA: Academic Press.

Metcalfe, T.S. (2005). The rise and citation impact of astro-ph in major journals. Bull. Am. Astron. Soc. 37, 555-557, arXiv:astro-ph/0503519v1

Metcalfe, T.S. (2006). The citation impact of digital preprints archives for solar physics papers. Solar Physics 239(1-2), 549-553, arXiv: astro-ph/0607079v2

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.

Moed, H. F.; Luwel, M. and Nederhof, A. J. (2002). Towards Research Performance in the Humanities. Library Trends, 50(3), 498-520.

Nederhof, A. J. and Noyons, E. C. M. (1992). International comparison of departments' research performance in the humanities. Journal of the American Society for Information Science 43(3): 249-256.

Nederhof, A.J., and Van Raan, A.F.J. (1993). A Bibliometric Analysis of Six Economics Research Groups: A Comparison with Peer Review. Research Policy 22: 353-368.

Neri, F. and Rodgers, J.R. (2006). Ranking Australian economics departments by research productivity. Economic Record, 82(1), 74-84.

Nicolaisen, J. (2007). Citation analysis. Annual Review of Information Science and Technology, 41, 609-641.

Norris, M.; Oppenheim, C. and Rowland, F. (2008). The citation advantage of open-access articles. Journal of the American Society for Information Science and Technology, 59(12), 1963-1972.

Pelzer, N. L. and Wiese, W. H. (2003). Bibliometric study of grey literature in core veterinary medical journals. Journal of the Medical Association, 91(4), 434–441.

Perneger, T. V. (2004). Relation between online "hit counts" and subsequent citations: prospective study of research papers in the BMJ. BMJ, 329, 546-547.

Pierce, S. J. (1992). On the Origin and meaning of bibliometric Indicators: Journals in the Social Sciences. Journal of the American Society for Information Science, 43(7), 477-487.

Robinson, W. C. and Poston, P.E. (2004). Literature Use by U.S. Economists Published in 1999,'' Behavioral & Social Sciences Librarian 22(2), 53–65.

Rodgers, J. R. and Neri, F. (2007). Research productivity of Australian academic economists: Human-capital and fixed effects. Australian Economic Papers, 46(1), 67-87.

Salman, M.D.; Dewell, R.; Willeberg, P.; GrandMaison, N.; Schoenbaum, M. and Moothart, T. (2007). Citations of plenary-session articles published in the Proceedings of the International Symposium on Veterinary Epidemiology and Economics. Preventive Veterinary Medicine, 80(1), 82-86.

Schwarz, G., and Kennicutt, R. (2004). Demographic and citation trends in astrophysical papers and preprints. Retrieved September 19, 2008, from http://arxiv.org/abs/astro-ph/0411275

Søndergaard, T.F., Andersen, J. and Hjørland, B. (2003). Documents and the communication of scientific and scholarly information, Revising and updating the UNISIST model. Journal of Documentation, 59(3), 278-320.

UNISIST (1971). Study Report on the feasibility of a World Science Information System, By the United Nations Educational, Scientific and Cultural Organization and the International Council of Scientific Unions. Paris, UNESCO.

Van der Meulen B. and Leydesdorff L. 1991. Has the Study of Philosophy at Dutch Universities Changed Under Economic and Political Pressures? Science, Technology & Human Values. 16(3): 288-321.

Van Raan, A.F.J. (2004). Measuring Science. Capita Selecta of Current Main Issues. In: H.F. Moed, W, Glänzel and U.Schmoch, editors: Handbook of Quantitative Science and Technology Research, Dordrecht: Kluwer Academic Publishers, 2004, p.19-50

Walker, R. D. & Hurt, C. D. (1990). Scientific and Technical Literature. An introduction to forms of communication. Chicago and London: American Library Association.

Whitley, R. (1991). The organisation and role of journals in economics and other scientific fields. Working Paper 204. Manchester business School.

Zhang, L. (2007). Discovering information use in agricultural economics: A citation study. The Journal of Academic Librarianship, 33(3), 403-413.

Appendix 1. High impact journals

	Average number of publications	Average number of citations	Average number of citations excluding self-citations	Average JIF	Average JIF excluding journal self-citations
Ecological Economics	101.3	287	195.1	3.07	1.86
Econometrica	59.4	272.4	280	5.19	4.71
Economic Geography	19.1	76.8	79.3	4.58	4.18
Economic Journal	77.9	232.2	236	3.34	3.05
Health Economics	65.4	270.2	252.6	4.70	3.88
International Economic Review	49.1	88	84.4	2.02	1.71
Journal of Development Economics	67.2	104.3	106	1.73	1.56
Journal of Econometrics	86.1	236.7	235.1	3.19	2.77
Journal of Economic Literature	17.9	209.5	219.7	12.80	12.35
Journal of Economic Theory	95.1	181.9	154.2	2.13	1.64
Journal of Economics & Management Strategy	24	38	38.5	1.78	1.56
Journal of Environmental Economics and Management	48.4	124.7	128.8	2.92	2.61
Journal of Industrial Economics	24.2	46.9	47.1	2.09	1.98
Journal of International Economics	55	174.6	173.7	3.53	3.11
Journal of Labor Economics	33.7	80.7	83.5	2.65	2.50

Appendix 2. Low impact journals

	Average number of publications	Average number of citations	Average number of citations excluding self-citations	Average JIF	Average JIF excluding journal self-citations
Bulletin of Indonesian Economic Studies	16.3	42.8	11	3.02	0.69
Canadian Journal of Economics	67	41.1	42	0.86	0.76
Contemporary Economic Policy	42.1	34.9	35.5	0.92	0.84
Economic Development and cultural Change	35.8	38.6	38.7	1.22	1.11
Economic Development Quarterly	27	34.5	21.8	1.36	0.82
Economica	34.2	34.8	38.8	1.19	1.15
Ekonomicky Casopis	58	17.3	2.4	0.34	0.04
Hitotsubashi Journal of Economics	10.2	1.9	1.2	0.24	0.12
Jahrbücher für Nationalökonomie und Statistik	43.1	7.7	3.5	0.20	0.08
Journal of Evolutionary Economics	23	26.8	23.6	1.26	1.00
Journal of Real Estate Finance and Economics	35.8	45.4	35.1	1.42	1.00
Oxford Bulletin of Economics and Statistics	37.5	49.2	50.8	1.53	1.31
Review of Income and Wealth	28	26.9	23.9	1.04	0.87
Scandinavian Journal of Economics	34.6	35.9	38	1.16	1.11
South African Journal of Economics	37.2	21.9	7.8	0.61	0.19

Appendix 3. Working papers

	Average number of publications	Average number of citations	Average IF	High impact working papers = 1
Boston College Working Papers in Economics	30.5	9.3	0.30	1
Cambridge Working Papers in Economics	37.2	14.2	0.38	1
Centre for Economic Policy Research, Research School of Social Sciences, Australian National University	16.9	3.6	0.21	0
Federal Reserve Bank of Minneapolis. Working papers	8	4.7	0.59	1
Harvard Institute of Economic Research Working Papers	3.5	1.92	0.55	1
School of Finance and Economics, University of Technology, Sydney	8.8	0.6	0.07	0
Universitat de Barcelona. Espai de Recerca en Economia. Working Papers in Economics	14.5	0.5	0.01	0
University of California at Los Angeles (UCLA). Economics Working papers	9.9	1.5	0.15	0
University of Copenhagen. Department of Economics. Discussion papers.	25.1	3.7	0.15	0
University of Rochester. RCER Working Papers.	10.30	4.4	0.43	1

12. Turning working papers into journal articles:

An exercise in micro-bibliometrics

Tove Faber Frandsen* & Paul Wouters**

* Corresponding author. Royal School of Library and Information Science, Birketinget 6, DK-2300 Copenhagen S, Denmark, E-mail: tff@db.dk

** Virtual Knowledge Studio for the Humanities and Social Sciences, Cruquiusweg 31, 1019 AT Amsterdam, The Netherlands, E-mail: paul.wouters@vks.knaw.nl

Abstract

This article focuses on the process of scientific and scholarly communication. Data on open access publications on the Internet not only provides a supplement to the traditional citation indexes but also enables analysis of the microprocesses and daily practices that constitute scientific communication. This article focuses on a stage in the life cycle of scientific and scholarly information that precedes the publication of formal research articles in the scientific and scholarly literature. Binomial logistic regression models are used to analyse precise mechanisms at work in the transformation of a working paper (WP) into a journal article (JA) in the field of economics. The study unveils a fine-grained process of adapting WPs to their new context as JAs by deleting and adding literature references, which perhaps can be best captured by the term sculpting.

Introduction

The process of scientific and scholarly communication is complex and comprises a variety of practices. This was recognized by the first pioneering studies of scientific communication in the field of psychology in the early 1960s (Garvey & Griffith, 1979; Garvey & Griffith, 1964). Communication processes are formed by many informal practices and processes that are difficult to capture in quantitative data. As a result, bibliometrics has mostly relied on data of the most formalized elements of the scientific and scholarly system of communication (Garvey, 1979), in particular, data pertaining to the formal scientific literature. This has resulted in a reliable body of knowledge about citation profiles and characteristics of journal sets, research groups, nations and individual researchers. Less knowledge is available, however, for other stages in the life cycle of scientific and scholarly information (Borgman, 2007). Although there have been a number of proposals to focus bibliometrics and informetrics more directly on the core of the creative process of knowledge making, such as context citation analysis (Cozzens, 1985) and citation analysis of concepts rather than of particular publications (Moed & Visser 2007), relatively little progress has been made in this respect.

In recent years, the rise of new media and the WorldWide Web (WWW) in communication practices of millions of people provides new opportunities for quantitative analysis of communication. This has created a number of new venues in research in both information science and bibliometrics. The most obvious is the creation of Web-based access to citation data, such as instantiated in the creation of the Web of Science (WoS). Although this does accommodate bibliometrics in some new ways, creating new problems for data quality management in the process, it does not entail conceptual innovation of the field. Second, the translation of citation analysis from document based to Web site-based studies (Almind & Ingwersen, 1997; Björneborn & Ingwersen, 2001; Rousseau, 1997) has resulted in the new field of Webometrics (Park & Thelwall, 2003; Thelwall, 2005; Scharnhorst & Wouters, 2006). Two recent approaches related to the field of Webometrics are especially promising. The first approach conceptualizes the Web as a complex self-organizing system and analyzes the development of its network characteristics on a much larger scale than traditional social network analysis (Albert & Barabasi, 2002; Newman, 2000;Wasserman, 1994). The

second approach is the design and analysis of new forms of visualization of massive, networked data sets (Börner, Chen, & Boyack, 2003).

The Web, however, not only is relevant to the study of large-scale aggregation of communication processes but also may open up new windows on the microprocesses and daily practices that constitute scientific communication. This study contributes to this line of research by zooming in on the process of research in more detail. We are interested in the process of writing that precedes the publication of formal research articles in the scientific and scholarly literature. In this respect, the use of the Web by researchers has basically opened up the area of what formerly was called "the grey literature" (Mili, 2000). Whereas working papers (WPs), preprints and other forms of grey literature were sometimes difficult to acquire in the past, they are now increasingly available on the Web, both in dedicated repositories and on the personal home pages of researchers (Kling, 2004). This has led to proposals for new business models for publishing scientific manuscripts, such as the "guild model" proposed by the late Rob Kling, which is built on peer reviewed series of WPs (Kling, Spector, & Mckim, 2002). Greater access to several forms of unpublished materials or preprints makes it possible to study the actual differences between WP and JA. Kling provides a precise definition of terminology regarding WP, preprints and e-prints, and critiques the general use of the term preprints for all sorts of grey literature. It follows the distinction between the "definitive publication" and all types of earlier versions of the work proposed by the International Working Group Report (1999–2000) of the International Association of STM Publishers (Kling, p. 595). Kling does not follow the Working Group's advice, however, to group all earlier versions together under the term "first publication." Instead, the term *manuscript* is proposed for documents that authors circulate prior to their acceptance for publication. The term *preprint* should then be used in a strict sense to refer to articles that have been accepted for a specific publication venue. In this terminology, the WP series that is the source for our data are venues for exchanging manuscripts. Because we restricted our analysis to those WPs that subsequently have been published, our data set consists of true preprints. On a related note, Kling also reminds us that increased exchange of manuscripts predates the Internet. For example, MIT's Research Laboratory of Electronics (RLE), its oldest and largest interdisciplinary research laboratory, has been issuing paper-based technical reports since it was founded in 1946 (Kling). Thus, "even in the paper-only world, publishing was a continuum" (Kling, p. 596).

In this article, we address the question, how do authors reformat their WPs to turn them into a formal JA? As far as we know, this question has not yet been studied in this way. It relates to earlier work in sociology on the acts of writing by Bazerman (1988), Knorr-Cetina and Harr (1981), and Latour and Woolgar (1986). However, this earlier work did not use citation analysis to throw light on the transformation of knowledge in the process of scientific communication. In the last few years, a small body of work has emerged in which conference proceedings have been used as data source for citation analysis. Drott (1995) called for more attention to the role of conference papers. Moed and Visser (2007) analyzed whether computer science conference proceedings can function as data source for evaluative bibliometrics. (For citation analysis in computer science, see also Goodrum, McCain, Lawrence, and Giles, 2001.) Recently, Montesi and Owen (2008) studied the transformation of conference presentations into formal JA in the field of software engineering and computer science, based on interviews with authors and journal editors. They conclude that "the topic of conference papers versus journal papers and of the process leading from one to the other still remains largely unexplored" (p. 816). In computer science, conference proceedings play a very important role in research communication. It is not known how many conference papers actually become journal papers in computer science. The median time lag between conference publication and journal publication in computer science was 2–4 years, which may discourage many authors from turning their manuscripts into JAs. The transformation of conference paper into journal article in this area is mainly a process of extension, and sometimes merging (Montesi & Owen, 2008, p. 826). This relates to the different roles played by conferences and journals. Montesi and Owen (2008) conclude that the central idea or problem of a particular work tends to remain unchanged. However, the journal article puts the work more definitively in the context of the discipline. They conclude: "Conference publication could be seen as innovation-laden, and journal publication as intended to settle a knowledge basis" (p. 827). Reference lists play an important role in this embedding and contextualization of knowledge. Although Montesi and Owen (2008) interviewed authors and editors about the different components of the scientific article that underwent change, they did not explicitly pose questions about references and citations. In contrast and complementary to this research, we have studied the characteristics of the reference lists of WPs and JAs. How do authors change the references in their papers, both in terms of deleting and adding

references? Which characteristics of authors, references and publications are influential in this process of pruning and adding?

These questions are also relevant to the more general concern of open access to scientific research (Bachrach et al., 1998; Ginsparg, 1997; Harnad, 1995). Open access-based resources are being used as pools of documents for citation analysis as an alternative to the citation databases traditionally used (Zhao, 2005). Reproducing the pool of open access documents in CiteSeer used by Zhao has revealed that a vast majority of the publications are grey literature (primarily conference papers and WPs).[24] A similar tool for citation analysis is CitEc, which uses data from Research Papers in Economics (RePEC) as the pool of documents. RePEC aims to enhance the dissemination of research in economics and consists of a decentralized database of WPs, JAs, and software components. CitEc is based on a selection of the publications in RePEC as the processing is not yet completed (approximately 25%). For each document made available in electronic format, the list of references is automatically indexed using CiteSeer algorithms in the process of identification and parsing of references.

Presently, the bulk of open access is WPs and preprints rather than publications in proprietary scientific journals. Consequently, when using open access-based resources for citation analysis, it seems important to know more about the citation characteristics of the different forms of publication, ranging from first drafts available on the Web and conference presentations to more structured WP and the final results in scientific and scholarly journals. In this article, we report on the last stage in this cycle: the transformation from structured working paper to journal article.

[24] Zhao (2005) conducts searches in December 2001 using the search terms "XML" and "extensible Markup Language" in header fields to identify publications on XML. Identical searches are conducted January 2008 and thus the results are not necessarily entirely identical.

Methods

The link between a journal article and a preceding version of the publication can be established relatively straightforwardly within those disciplines that are using unrefereed early versions of publications as a means to communicate research. Some disciplines have well-established traditions of paper manuscript publication going back to the 1960s. Examples of these are artificial intelligence, computer science, economics, demography, linguistics, and high-energy physics. The number of disciplines including this practice has subsequently increased to also include information systems, physics, and mathematics (Kling et al., 2002). The level of support within a field of publishing manuscript papers is highly dependent on the specific characteristics of the field (Kling & McKim, 1999, 2000). Economics is such a discipline. It is characterized as a pronounced journal publishing discipline in which WP also play a central role (Pierce, 1992; Whitley, 1991).An economics WP are often later published as JA (Carraro, Lanza & Papponetti, 2003). Because they expect it will lead them to double counting, Goyal, van der Leij, and Moraga-Gonzalez (2006) exclude WP from their analysis of the publishing behavior of economists, even though they make use of a database (EconLit) that includes WP.

The analyses in our study are based on a sample of WPs subsequently published as JAs. A 5% sample was drawn randomly among the 15,500 author contact and publication listings available on RePEc (www.repec.org). The sample was drawn at the level of the first letter in the surname of the authors. Consequently, some authors are present in the dataset with multiple publications. In RePEC, individual authors can create and maintain a profile of their work. The list of publications by an author separates WPs from JAs, but the author can create a link between a WP and a JA if the JA is a revised version of the WP. Publications were excluded from this study if (a) the author did not include WPs transformed into JAs in their profile or (b) relevant information linking the two publications was lacking. Furthermore, the publications were excluded from the study if the WP was published in a subscription-based WP series (e.g., NBER Working Papers) and if it was not available by open access. Finally, publications were excluded if either the WP or the JA was published before 2000 (to avoid bias as there were very few early publications). The data collection resulted in a set consisting of 135WPs and 135 matching JAs.

The data was analysed in two separate models, and in both analyses, the dependent variable had only two possible outcomes: 0 and 1. In the first model, 0 represents references that were present both in WP and in JA, and 1 represents references that were present only in the WP. Correspondingly, in the second model, 0 represents references that were present both in WP and in JA and, 1 represents references that were present only in the JA.A reference changing status in the time passing from WP to JA is not given the value of 1, as it essentially is the same publication. A reference can change status from, for example, unpublished manuscript to published journal article or from WP to JA. The link between the potential publications was established using RePEC, and in case of missing information on RePEC, the authors' personal Web sites were consulted. Binomial or binary logistic regression is a form of regression which is used when the dependent variable is a dichotomy and the independents are of any type. Logistic regression applies maximum likelihood estimation after transforming the dependent variable into a logit variable (the natural log of the odds of the dependent occurring or not). In this way, logistic regression estimates the probability that a certain event will occur. Logistic regression coefficients can be used to estimate odds ratios for each of the independent variables in the model.

Consequently, a binary logit model was estimated to predict the publications most likely to undergo highest degree of revision and the references most likely to be revised as the WP was transformed into a JA. The extent to which references were excluded in the transition from working paper to journal article is probably influenced by many factors (such as characteristics of the author of the publication, characteristics of the reference, and characteristics of the journal in which the WP is later published). Although a number of variables were included, not all aspects of these factors could be controlled for in this study.

The explanatory variables may be grouped into the following two categories: publication and reference characteristics. The characteristics of the publications were described by two variables. First, the length of the period from WP to JA is included as a variable. As the length of that period increases, we expect more changes in the reference list because more relevant publications may have been published. The second variable describing the publication is a binary variable, indicating whether the journal is a core journal within the field. A core journal is defined as being 1 of the 27 journals on the list of core economics journals identified by Diamond (1989), which is available in

Appendix A. Impact factors could be used as an alternative approach to capture the effect of prestige or quality of the journal or publisher of the reference. However, not all publications in the study were source journals in the ISI citation databases[25], and, consequently, reasonable comparisons between source and non-source items seem unobtainable.

The references were characterised by a number of variables. First, the age of the reference relative to that of the JA was included as a variable to capture whether excluded or included references tend to be younger or older than the references present in both versions. Second, author and journal self-citations were included as variables as authors may choose to include more references in the final version to their previous publications (to make these more visible) or to the journal they are submitting it to (to increase the chances of acceptance). Third, un-published references were given the value of 1 in a variable to capture the effect of authors possibly excluding references without a publisher. Fourth, the number of citations of the first author of the reference was included in a variable because in the process of revising the WP authors may choose to include more references by highly cited authors and fewer references by authors with few citations.

The number of citations was determined using the Social Sciences Citation Index (SSCI). More precisely, the variable was determined as the number of citations of the first author of a given reference at the time of publication of the JA. Furthermore, an additional variable described if the author of a reference was present with more than one publication in the reference list. It is possible that the process of revising aWP includes that the author looks for more recent publications by authors already cited in WP and any effect of it can be detected by adding this variable to the analysis. Finally, a variable described whether the reference is published in an economics core journal.

[25] In calculating impact, the publications do not necessarily have to be included in the pool of citing documents covered by ISI. It is possible to construct impact factors for journals not covered by ISI (Stegmann, 1999; Stegmann & Grohmann, 2001). However, in the case of journals the relative size of journal self-citations varies on the basis of journal characteristics (Frandsen, 2007), and journal self-citations cannot easily be included in the calculation of impact of the nonsource journals. The case of monographs is no less complicated as the citing behaviour in monographs differs from journals (Hicks, 2004).

The following binomial logit model is the formal model for the multivariate analysis:

RRD = *f(*RA,RSC,RJSC,RU,RAC,RC,RD, PC, PP*)*

where:

RR_D (dependent variable) is revised reference. If the reference is present both in WP and in JA, RR_D =0, otherwise RR_D =1.

R_A is the age of reference relative to JA. It is measured as the publication year of the JA: the publication year of the reference +1.

R_{SC} is self-citation reference. If the reference is written by one or more of the authors of the publication, R_{SC} =1, otherwise R_{SC} =0.

R_{JSC} is journal self-citation reference. If the reference is published in the same journal as the publication, R_{JSC} =1, otherwise R_{JSC} =0.

R_U is un-published reference. If the reference is not a WP, JA, or monograph with a publisher, R_U =1, otherwise R_U =0.

R_{AC} is the number of citations of the first author of the reference.

R_{AP} is author of reference already present in the reference list. If the author is already represented in the reference list, R_{AP} =1, otherwise R_{AP} =0.

R_C is the reference published in an economics core journal. If the reference is published in an economics core journal, R_C =1, otherwise R_C =0.

P_C is the publication published in an economics core journal. If the publication is published in an economics core journal, P_C =1, otherwise P_C =0.

P_P is the maximum number of years from publication as WP to publication in journal, measured as the publication year of the JA: the publication year of the WP +1.

The model was applied separately to the two data sets (exclusion and inclusion of references). For each of the analyses, the sample included the references from all 135 publications (revised and unrevised). Consequently, the samples also included publications with reference lists identical in the WP and subsequent JA.

Results

The reference list in a WP that was subsequently published as a JA comprises a number of references included and excluded in the subsequent JA. On average, 15.4% of the references is being excluded for the JA. Furthermore, the reference list in a JA that was initially a WP comprises a number of references from the WP and a number of new references. The average JA (that was formerly a WP) includes 15.5% of the references that were not present in the WP. Some WPs undergo more revision than others and the averages of 15.4% and 15.5% are calculated on the basis of great differences. The former average is calculated on the basis of percentages ranging from 0 to 69.0 and the latter on the basis of percentages ranging from 0 to 71.0.As a result, some of the reference lists in the 135 publications undergo no revision (13.3%), some JAs have no added references (20.0%), and some WPs have no references being omitted in the JA (32.6%). Consequently, a majority of WPs undergo revision, and adding references occurs more often than omitting references. Figure 1 is an illustration of the correlation between the share of references in a WP being omitted and the share of references in the JA that is added after publishing the WP.

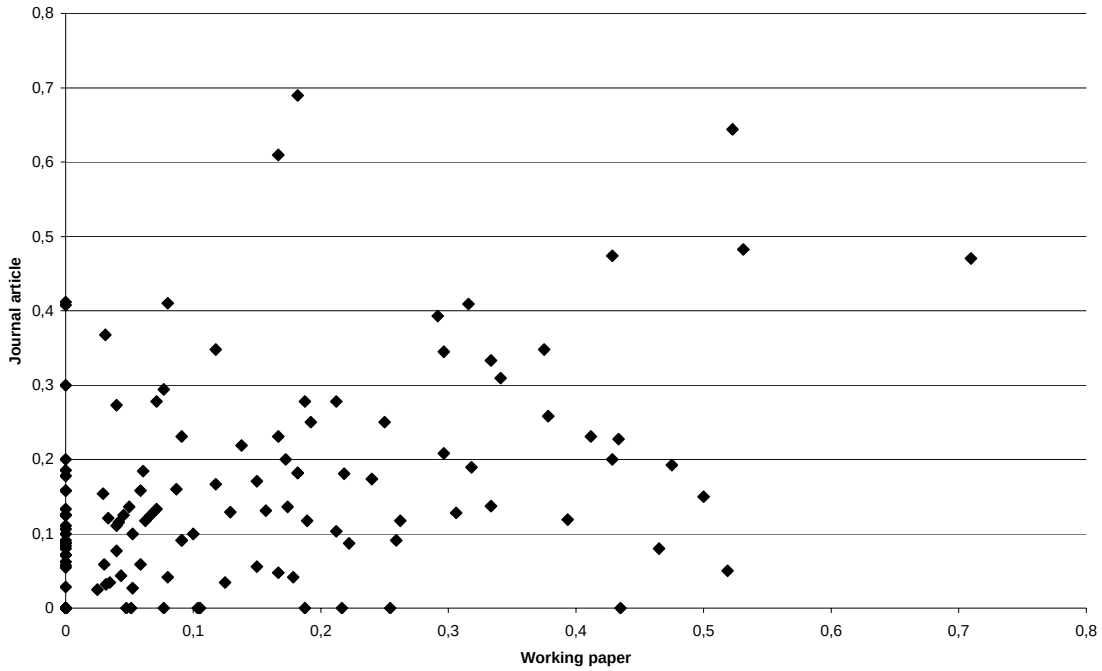


Figure 1. Share of references being added to the JA and omitted from WP.

Each dot is a linked JA/WP pair that is to be interpreted as follows: A pair with a value of .7 on the x axis and .5 on the y axis is a WP in which 70% of the references are not present in the subsequent JA, and 50% of the references in the JA are not present in the WP. The degree of adding and omitting references is positively correlated (statistically significant at the .01 level). However, the linear relationship is weak (an R2 of .21 indicates that we are able to explain only 21% of the variation in adding of references by the omission of references).

The reference list in a WP is often undergoing considerable revision prior to publication in a journal. However, some WPs undergo more revision than others and some references are more likely to be excluded or included. The following tables provide information on the characteristics of the document or the reference increasing or decreasing the likeliness of revision.

The interpretation of a logistic regression coefficient is not as straightforward as that of a linear regression coefficient. Because the logit transformation is nonlinear, it does not mean a constant increase in probability. Consequently, in this case, the results are

presented as just the signs of the coefficients (the coefficients are available in Appendixes A and B), followed by an analysis of the effect of the statistically significant variables. The common measures of association for logit analysis (probabilities, odds, or their ratios) are easily confused; however, probability pairs are easily interpreted (Liberman, 2005). Overviews of the dependent and independent variables in the two regressions are available in Appendixes B and C. Goodness-of-fit of the two models is measured by a likelihood ratio test and the model of exclusion of references is significant at the .01 level according to the model chi-square statistic (- 2 log likelihood relative to null model: 222.47 with 10 degrees of freedom).[26] The model of inclusion of references is significant at the .01. level according to the model chi-square statistic (- 2 log likelihood relative to null model: 205.93 with 10 df). To diagnose potential multicollinearity, the correlations among the variables were examined and none of them were larger than .26 for the first model and .28 for the second model. All correlations are available in Appendixes D and E. Furthermore, no symptoms were found of multicollinearity (such as excessive logit iterations or inflated standard errors), which indicates that multicollinearity should not cause disturbance in this case. Un-published self-citations were added to the models as interaction variable because it contributed significantly in one of the models.

Table 1 provides information on the characteristics of references that is coming in or being omitted from the WP as it becomes a JA.A WP becoming a JA will have more excluded references if the journal is an economics core journal or the time from WP to JA is relatively long. The references being excluded are more likely to be unpublished material or by an author already present in the reference list. On the other hand, excluded references are less likely to be publications from authors with many citations, published in the same journal (a journal self-citation), or published in core economics journals. Finally, excluded references are more likely to be unpublished self-citations and less likely to be self-citations in general.

[26] In the linear regression model R^2 summarizes the proportion of variance in the dependent variable associated with the independent variables. Large R^2 values indicate that more of the variation is explained by the model, up to a maximum of 1. For regression models with a categorical dependent variable, it is not possible to compute a single R^2 statistic that has all of the characteristics of R^2 in the linear regression model, so pseudo-R^2 are computed instead (Powers & Xie, 2000).

	References omitted		**References added**	
	Sign of coeff.	**P-value**	**Sign of coeff.**	**P-value**
Age of reference (relative to JA)		.33	-	.00
Reference is an author self-citation	-	.03		.72
Reference is a journal self-citation	-	.02		.86
Reference is un-published document	+	.00		.59
Citations of first author of reference	-	.00	-	.09 *
Reference is by author already cited in the document	+	.00	+	.00
Reference is published in an economics core journal	-	.04		.01
The journal is an economics core journal	+	.00		.71
Number of years from WP to JA	+	.00	+	.00
Unpublished self-citations	+	.02		.57
# of observations	3963		3968	

Table 1. Binary logistic regression.

Note. WP, working papers; JA, journal articles. Dependent variable: references present both in WP and in JA (0) or present only in WP/JA (1). Only the signs of the statistically significant variables are shown. Variables are significant at the .05 level.
 Variables significant at the .1 level.

Only the signs of the statistically significant variables are shown (variables are significant at the .05 level and variables significant at the .1 level are marked with *). However, all signs are available in appendix.

Furthermore, Table 1 shows that a WP becoming a JA will have more added references if the journal is an economics core journal or if the time from WP to JA is relatively long. The included references are more likely to be by an author already present in the reference list. Furthermore, included references are characterized by being relatively younger. They are also more often self-citations. This is expected because some of the included references may have been published after the WP was finished. Finally, included references are more likely to be written by authors with few citations.

To better understand the size of the effects coming from the different variables, changes in probabilities when changing the characteristics of an article, a journal or the author can be computed. Table 2 presents the results of such computations. The information needed for the calculation is available in Appendixes F and G.

	References coming in	**References omitted**
Probability computed at mean (per cent)	13.4	11.2
Reference is 2 years younger	0.9 *	0
Reference is an author self-citation	0	-3.7 **
Reference is a journal self-citation	0	-6.4 *
Reference is unpublished	0	26.1 *
Author of reference has 500 more citations	-0.3 *	-0.4 *
Reference is from a core journal	-2.8 *	-2.0 *
Article is published in a core journal	0	3.5 *
2 more years pass before article is published	8.8 *	3.8 *
Reference is by author already cited	10.9 *	10.2 *
Reference is un-published self-citation	0	15.6 *

Table 2. Quantitative effects of changing characteristics in percentage points. *Note*. The estimated probability in the first row of the table is calculated at mean values for the numbered variables and at the value zero for the binary variables. The changes in probability are the change in the estimated probability when changing the characteristics one at a time.

□Variables significant at the .05 level.

□□Variables significant at the .1 level.

A first glance at Table 2, reveals that some of the variables are much more important than others. If the author of reference has 500 more citations, for instance, there is only a small change in the estimated probability of it being an added reference (of −0.3 percentage points). On the other hand, an unpublished document has a more than 25 percentage points higher probability of being omitted than a published article with the

same background characteristics. An unpublished document authored by one or more of the same authors has a more than 15 percentage points higher probability of being excluded. We were surprised to learn that the age of the reference means relatively little for the estimated probability. Thus, an article that is two years younger has a probability of being added that is a .9 percentage point higher than an article of mean age. It is striking that the characteristic indicating whether the author of a reference has already been cited in the document has such a large impact. If an author had already been cited in the document, then this raises the probability both of being added and of being omitted by approximately 10 percentage points. The number of years passing from the WP to the final publication of the JA has a large impact on the adding of new references. There is a 9 percentage points higher probability for a given entry in the reference list of being an added one if the time span between WP and article is two years longer than would have been the case for an otherwise identical reference in a paper with just the mean time span between working paper and article.

Discussion

Montesi and Owen (2008) conceptualize the transformation from conference paper to journal article in software engineering as primarily a process of extension, and sometimes of merging. The economics working paper bears great resemblance with early JAs (Zhou, 2001) and is typically longer than the journal article. Consequently, contraction could be an equivalent conceptualization of the process from working paper to journal article. Our results indicate that this may be too limited. Rather than reducing only the WP, we see a process both of deleting and of adding references. Basically, we are witnessing a process in which the author tries to maximize the fit between the manuscript and the scholarly journal in which she wants the paper published. In this process, deleting and adding references is a very important and telling activity. The probability that a reference will be deleted is influenced by characteristics both of the working paper and of the reference. References to unpublished material have a high chance of being deleted (25 percentage points higher than average). However, in the case of references to unpublished work that are by one or more of the authors the risk of

being deleted is noticeably higher (41 percentage points higher than average).[27] Apparently, researchers are generally reluctant to include unpublished documents in their own work, although grey literature can be a vital resource for literature reviews (Benzies, Premji, Hayden, & Serrett, 2006; Hopewell, McDonald, Clarke, & Egger, 2007). Even though the deletion of unpublished materials is as expected, it has an important implication. An unpublished document will generally be cited in a paper only if it has actually been used in the research process. Its deletion means that scientific and scholarly journals tend to deemphasize the communicative role of (yet) unpublished documents. In other words, grey literature is made less visible in citation analysis of the formal literature. This means that WPs more completely represent the actual reading that went into the research than the journal article. Therefore, if one wants to study communication processes and processes of information transfer in research, citation analysis of WPs may be more fruitful than the traditional citation analysis of JAs. The extent to which this should be a priority depends, of course, on the role WPs play in the process of scientific communication. Citation analysis of WPs will face additional technical problems and it can be expected that either it will have to be conducted manually or at least the results will have to be verified manually.[28]

References to works by one or more of the authors have a lower risk of being deleted (about 4 percentage points less than average). As stated, this is just the opposite in the case of unpublished self-citations. The results indicate that authors are reluctant to remove references by themselves. A possible explanation is that self-citations serve necessary functions. According to Gami, Montori, Wilczynski, and Haynes (2004), self-citations allow an author or a group to expand on previous hypotheses, refer to established study designs and methods, or justify further investigations on the basis of prior results. However, the results do not indicate that authors deliberately add references to their own publications merely to self-cite and thus perhaps artificially inflate an article's importance to the general scientific community as suggested by Hyland (2003) and satirized by Craddock, O'Donovan, and Owen (1996).

[27] This additional analysis is done by estimating the risk for the three possible combinations of unpublished documents and self-citations separately.

[28] We can expect that citation analysis of working papers will face the same kind of technical problems as citation analysis of conference proceedings. The latter are outlined in (Moed & Visser, 2007).

We were somewhat surprised to learn that the effect of high impact authors is far less than one would expect given the general emphasis on impact factors. This may mean that the composition of the reference list is mostly influenced by considerations relating to the particular paper rather than to more general considerations of citing influential authors. Apparently, symbolic citing behavior (citing for the sake of citing famous authors) does not have a big impact on the results in our study.

Our results indicate that authors do actively follow the developments in the published literature. If it takes 2 more years before the journal article is published, then the chance that references are added to the paper increases with 11 percentage points. On the other hand, this process is still highly selective. If a reference is 2 years younger than average, its chance of being added to the working paper increases only with .9 percentage points.

Our results also indicate how authors try to increase the "fit" between the manuscript and the journal. If the reference is a journal self-citation, it has a lower chance of being deleted (5 percentage points less than average), but it does not increase its chance of being added. Apparently, this process of adaptation of the working paper manuscript to the existing literature is asymmetrical. This may have different causes at the level of the individual author, which probably interact to give this aggregate result at the level of representative samples of articles. Authors may have already checked all relevant references in the journal and see therefore no reason to add references. Or they may simply restrict themselves, in general, to the already existing list of references in the working paper and only add references from novel work whether or not it is published in the target journal. The result does, again, point to the limited role of purely symbolic citation behavior.

Authors who have already been cited in the working paper have a higher chance of having publications both deleted and added. It is not clear from our study what the mechanisms are behind this outcome. An obvious interpretation may be that for every single article there is a restricted number of authors working within that field, and, consequently, the process of adding and deleting references takes place within that relatively short list of authors. Within the time period from the WP to JA, some of the authors on that list could have published additional publications within the field leading to the removal of old references and the addition of new references. However, it appears that authors tend to keep the number of references by a given author in the reference list

relatively stable. Authors could also choose to keep the references from the WP and just add more, recent references from authors already present in the reference list. One possible explanation might be that authors want to condense their list of reference and, as a result, decrease the variety within their list while, at the same time, duplicating the list. This is in need of further studies, perhaps in a research design combining citation analysis with interviews.

This study has a number of limitations that are worth reviewing before concluding on the findings. First, and most important, is the limitation due to the sample size. The analyses are based on 135WPsand a corresponding number of JAs. Analyses on the level of single references typically involve relatively small publication sample; therefore, the results should be generalized with great caution. Second, the sample of WPs may not be representative of all economics papers, as they are drawn from author profiles on RePEC and not all economics researchers create profiles.

Conclusion

This study has unveiled a fine-grained process of adapting WP to their new context as journal article by deleting and adding literature references. This can perhaps be best captured by the term sculpting as this reflects that the core structure of the working paper is maintained while important changes are made to several of its components (Montesi & Owen 2008). The list of references is a very important building block of the scholarly article. Authors often start to check the references to see whether a particular article is of their interest. Because we can think of the process of publication of a paper as a process of embedding one's ideas and findings into the body of existing work, it is not least the reference list in relation to the text as a whole that is instrumental in this embedding.

Rather than limiting ourselves to the quantitative description of references and citations, we have used binomial logistic regression models. This enabled us to estimate the magnitude of the effects of attributes on the process of addition or deletion, at the level of both the individual reference and the working paper and the journal article. Our study shows that applied in this way, citation analysis, or more correctly reference analysis, is able to reveal the subtleties of this creative process in ways that interviews and surveys

will never be able to. Perhaps it is not coincidental that Montesi and Owen (2008) did not include the reference list in their survey. Authors tend to forget what they actually did with particular references. Still, these "forgotten acts" may be crucial in shaping the body of scholarly work because they create the citation network structure of the formal literature.

As we have indicated in the discussion of our results, some of our findings are in need of more detailed explanation. They may be idiosyncratic to our particular data set, or they may refer to more general processes that are also operative in other fields of research or scientific journals. Perhaps the two most important limitation of our study are as follows: (a) the variation of the relationship between publication and communication venues in different fields of the sciences (including humanities) and (b) the different factors influencing the decision processes with respect to the reference lists. Because we have studied only a limited sample in one field, extension of this work to other fields may reveal quite different patterns. It seems worthwhile to pursue analysis of other forms of communication around manuscripts in a variety of fields and compare how these are reflected (or hidden) in different forms of citation analysis. Second, it is clear that referees and editors do influence the ways authors add and delete references to their manuscripts. However, we do not know how important this influence is, nor how it relates to the factors we were able to discern in this study. Journals vary greatly in their instructions to authors (Schriger, Arora, & Altman, 2006). We have anecdotal evidence that sometimes editors may try to enhance the impact factor of their journal by advising authors to include more journal self-citations (Falagas & Alexiou, 2007, 2008). However, we have no systematic knowledge of the role of editors or referees in the shaping of the reference lists of scholarly publications.
(Smith, 1990) does not discuss the role of the editor or referee in suggesting or influencing the reference lists.

To sum up, we propose that a better understanding of the precise mechanisms at work in the transformation of WPs into JAs is highly relevant in a scholarly universe where open access to WPs seems to increasingly shape and transform the reading habits of working scientists and scholars, perhaps already more so than access to the formal journal literature. We hope that our exercise in microbibliometrics and the modelling approach applied may encourage further research within this area.

Acknowledgments

We are grateful to Birger Hjørland, Jan Kok, Jeppe Nicolaisen, Ernst Thoutenhoofd, and Sally Wyatt for their insightful comments about this article.

References

Albert, R., & Barabasi, A.L. (2002). Statistical mechanics of complex networks. Reviews of Modern Physics, 74(1), 47–97.

Almind, T., & Ingwersen, P. (1997). Informetric analyses on the World Wide Web: Methodological approaches to "webometrics". Journal of Documentation, 53(4), 404–426.

Axarloglou, K. & Theoharakis, V. (2003). Diversity in economics: An analysis of journal quality perceptions. Journal of the European Economic Association, 1(6), 1402–1423.

Bachrach, S., Berry, R.S., Blume, M., Foerster, T.v., Fowler, A., Ginsparg, P., et al. (1998). Intellectual Property: Who should own scientific papers? Science, 281(5382), 1459–1460.

Bazerman, C. (1988). Shaping written knowledge: The genre and activity of the experimental article in science. Madison: University of Wisconsin Press.

Benzies, K.M., Premji, S., Hayden, K.A., & Serrett, K. (2006). State-of-theevidence reviews: Advantages and challenges of including grey literature. Worldviews on Evidence-based Nursing, 3(2), 55–61.

Björneborn, L., & Ingwersen, P. (2001). Perspectives of Webometrics. Scientometrics, 50(1), 65–82.

Borgman, C. (2007). Scholarship in the digital age. Cambridge, MA: MIT Press.

Börner, K., Chen, C., & Boyack, K.W. (2003) in Annual Review of Information Science and Technology, ed. Cronin, B. (Information Today, Inc., Medford, NJ), Vol. 37, 179–255.

Bräuninger, M.,&Haucap, J. (2003).What economists think of their journals and how they use them: Reputation and relevance of economics journals. Kyklos, 56(2), 175–197.

Burton, M.P., & Phimister, E. (1995). Core journals: A reappraisal of the Diamond list. The Economic Journal, 105(429), 361–373.

Carraro, C., Lanza, A., & Papponetti, V. (2003). One thousand working papers. Fondazione Eni Enrico Mattei, Working Papers: 2003.1000. Retrieved August 2008 from http://www.feem.it/NR/rdonlyres/7F4076E6-1D40-4185-B33B-92F032206643/947/No tamilleultimissssssima.pdf

Cozzens, S.E. (1985). Comparing the sciences: Citation context analysis of papers from neuropharmacology and the sociology of science. Social Studies of Science, 15, 127–153.

Craddock, N., O'Donovan, M.C.,&Owen, M.J. (1996). Introducing Selfcite 2.0—Career enhancing software. British Medical Journal, 313, 1659–1660.

Diamond, A.M. (1989). The core journals of economics. Current Contents, 1, 4–11.

Drott, M.C. (1995). Reexamining the role of conference papers in scholarly communication. Journal of theAmerican Society for Information Science, 46(4), 299–305.

Falagas, M.E., & Alexiou, V.G. (2007). Editors may inappropriately influence authors' decisions regarding selection of references in scientific articles. International Journal of Impotence Research: Official Journal of the International Society for Impotence Research, 19(5), 443–445.

Falagas, M., & Alexiou, V. (2008). The top-ten in journal impact factor manipulation. Archivum Immunologiae Et Therapiae Experimentalis, 56(4), 223–226.

Frandsen,T.F. (2007). Journal self-citations—Analysing the JIF mechanism. Journal of Informetrics, 1(1), 47–58.

Gami,A.S., Montori,V.M.,Wilczynski, N.L.,&Haynes, R.B. (2004).Author self-citation in the diabetes literature. Canadian medical Association Journal, 170(13), 1925–1927.

Garvey, W.D., & Griffith, B.C. (1964). Scientific information exchange in psychology. Science, 146, 1655–1659.

Garvey,W.D., & Griffith, B.C. (1979). Scientific communication as a social system. In W. Garvey D. (Ed.), Communication: The essence of science (pp. 148–164). Oxford: Pergamon Press.

Garvey, W.D. (1979). Communication: The essence of science. Oxford: Pergamon Press.

Ginsparg, P. (1997). Winners and losers in the global research village. The Serials Librarian, 1997, 30(3–4), 83–95.

Goodrum, A.A., McCain, K.W., Lawrence, S., & Giles, C.L. (2001). Scholarly publishing in the internet age: A citation analysis of computer science literature. Information Processing & Management, 37(5), 661–676.

Goyal, S., van der Leij, M., & Moraga-Gonzalez, J.L. (2006). Economics: An emerging smallworld. Journal of Political Economy, 114(2), 403–432.

Harnad, S. (1995). Implementing peer review on the net: Scientific quality control in scholarly electronic journals. In R. Peek & G. Newby (Eds.), Scholarly publishing: The electronic frontier. Cambridge,MA:MITPress.

Hicks, D. (2004). The four literatures of social science: Handbook of quantitative science and technology studies, Henk Moed (Ed.),

Hopewell, S., McDonald, S., Clarke, M., & Egger, M. (2004). Grey literature in meta-analyses of randomized trials of health care interventions. Cochrane Database of Systematic Reviews, 18(2), art. no.: MR000010.

Hyland, K. (2003). Self-citation and self-reference: credibility and promotion in academic publication. Journal of the American Society for Information Science and Technology, 54(3), 251–259.

Kalaitzidakis, P., Mamuneas, T.P., & Stengos, T. (1999). European economics: An analysis based on publications in the core journals. European Economic Review, 43(4–6), 1150–1168.

Kling, R. (2004). The Internet and unrefereed scholarly publishing. Annual Review of Information Science and Technology, 38, 591–631.

Kling, R.&McKim, G. (1999). Scholarly communication and the continuum of electronic publishing. Journal of the American Society for Information Science, 50(10), 890–906.

Kling, R. & McKim, G. (2000). Not just a matter of time: Field differences and the shaping of electronic media in supporting scientific communication. Journal of the American Society for Information Science, 51(14), 1306–1320.

Kling, R., Spector, L., & Mckim, G. (2002). Locally controlled scholarly publishing via the internet: The guild model. The Journal of Electronic Publishing, 8(1).

Knorr-Cetina, K.D., & Harr, R. (1981). The manufacture of knowledge: An essay on the constructivist and contextual nature of science. Oxford: Pergamon Press.

Latour, B., & Woolgar, S. (1986). Laboratory life: The construction of scientific facts (2nd ed.). Princeton University Press.

Liberman, A.M. (2005). How much more likely? The implications of odds ratios for probabilities. American Journal of Evaluation, 26(2), 253–266.

Medoff, M.H. (2006). The efficiency of self-citations in economics. Scientometrics, 69(1), 69–84.

Moed, H.F., & Visser, M.S. (2007). Developing bibliometric indicators of research performance in computer science: An exploratory study. Leiden: Leiden University.

Newman, M.E.J. (2000). The structure of scientific collaboration networks.Proceedings of the National Academy of Sciences, 98(2), 404–409.

Park, H., & Thelwall, M. (2003). Hyperlink analyses of the world wide web: A review. Journal of Computer Mediated Communication, 8(4).

Pierce, S.J. (1992). On the Origin and meaning of bibliometric Indicators: Journals in the Social Sciences. Journal of the American Society for Information Science, 43(7), 477–487.

Powers, D., & Xie, Y. (2000). Statistical methods for categorical data analysis. San Diego, CA: Academic Press.

Rousseau, R. (1997). Sitations: An exploratory study. Cybermetrics, 1(1). Available at: http://www.cindoc.csic.es/cybermetrics/articles/v1i1p1.html

Scharnhorst, A., & Wouters, P. (2006). Web indicators—A new generation of S&T indicators? Cybermetrics, 10(1), paper 6.

Schriger, D.L., Arora, S., & Altman, D.G. (2006). The content of medical journal instructions for authors [Abstract]. Annals of Emergency Medicine, 48(6), 743–749.

Smith, A.J. (1990). The task of the referee. IEEE Computer, (April), 65–71.

Stegmann, J. (1999). Building a list of journals with constructed impact factors. Journal of Documentation 55(3), 310–324.

Stegmann, J., & Grohmann, G. (2001). Citation Rates, Knowledge Export and International Visibility of Dermatology Journals listed and not listed in the Journal Citation Reports. Scientometrics 50(3), 483–502.

Sussmuth, B., Steininger, M., & Ghio, S. (2006). Towards a European economics of economics: Monitoring a decade of top research and providing some explanation. Scientometrics, 66(3), 579–612.

Sutter, M., & Kocher, M. (2001). Tools for evaluating research output: Are citation-based rankings of economics journals stable? Evaluation Review, 25(5), 555–566.

Thelwall, M. (2005). Link analysis: An information science approach. San Diego: Academic Press.

Whitley, R. (1991). The organisation and role of journals in economics and other scientific fields.Working Paper 204. Manchester business School.

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.

Zhou, J.Z. (2001). The evolution of journals and early scholarly publications from the ancient China to the 21st century. Journal of Information, Communication, and Library Science, 17(3), 1–11.

Appendix A. Diamond core journals

American Economic Review
Brookings Papers on Economic Activity
Canadian Journal of Economics
Econometrica
Economic Enquiry
Economic Journal
Economica
Economics Letters
European Economic Review
International Economic Review
Journal of Development Economics
Journal of Econometrics
Journal of Economic Literature
Journal of Economic Theory
Journal of Financial Economics
Journal of International Economics
Journal of Labor Economics
Journal of Law and Economics
Journal of Mathematical Economics
Journal of Monetary Economics
Journal of Political Economy
Journal of Public Economy
Oxford Economic Papers
Quarterly Journal of Economics
RAND Journal of Economics
Review of Economic Studies
Review of Economics and Statistics

Appendix B. The dependent and independent variables included in the binary logistic regression on exclusion of references

Variable	Range	Mean	Std. Dev.
Dependent variable			
Revised Reference	0 - 1	.15	.36
Independent variables			
Age of reference (relative to JA)	1 - 233	12.78	11.73
Reference is a self-citation	0 - 1	.07	.26
Reference is a journal self-citation	0 - 1	.05	.21
Reference is un-published document	0 - 1	.04	.20
Citations of first author of reference	0 – 17,802	1179.82	2153.92
Reference is by author already cited in the document	0 - 1	.18	.39
Reference is published in an economics core journal	0 - 1	.34	.47
The journal is an economics core journal	0 - 1	.41	.49
Number of years from WP to JA	1 – 8	3.01	1.23
Reference is an unpublished self-citation	0 – 1	.01	.09

Appendix C. The dependent and independent variables included in the binary logistic regression on inclusion of references[29]

Variable	Range	Mean	Std. Dev.
Dependent variable			
Revised Reference	0 - 1	.16	.36
Independent variables			
Age of reference (relative to JA)	1 - 233	12.35	11.55
Reference is a self-citation	0 - 1	.07	.26
Reference is a journal self-citation	0 - 1	.05	.22
Reference is un-published document	0 - 1	.03	.16
Citations of first author of reference	0 – 17,851	1173.18	2142.59
Reference is by author already cited in the document	0 - 1	.18	.39
Reference is published in an economics core journal	0 - 1	.34	.47
The journal is an economics core journal	0 - 1	.40	.49
Number of years from WP to JA	1 – 8	3.05	1.20
Reference is an unpublished self-citation	0 - 1	.004	.06

[29] It should be noted that the range, means and standard deviations of appendix 2 and 3 similar values which is due to the overlapping references of respectively 85 and 86 per cent of the data material.

Appendix D. Multi-collenearity among variables in the analysis of references being excluded when WP turns into JA

	R_A	R_{SC}	R_{JSC}	R_U	R_{AC}	R_C	P_C	P_P
R_{SC}	-.17**							
R_{JSC}	-.05**	.04**						
R_U	-.13**	.11**	-.05**					
R_{AC}	.28**	-.11**	-.05**	-.07**				
R_C	.06**	-.05*	.13**	-.14**	.10**			
P_C	.05**	.05**	.10**	-.04	.01	.11		
P_P	.10**	-.02	.01	.02	-.02	-.001	.02	
R_{AP}	-.09**	.26**	.03	.09**	.05**	-.02	.02	.01

**. Correlation is significant at the .01 level (2-tailed).

*. Correlation is significant at the .05 level (2-tailed).

Appendix E. Multi-collenearity among variables in the analysis of references being included when WP turns into JA

	R_A	R_{SC}	R_{JSC}	R_U	R_{AC}	R_C	P_C	P_P
R_{SC}	-.17**							
R_{JSC}	-.05**	.03*						
R_U	-.11**	.05*	-.04*					
R_{AC}	.30**	-.11**	-.05**	-.06**				
R_C	.07**	-.05**	.13**	-.10**	.11**			
P_C	.03	.06**	.09**	-.03	.01	.11**		
P_P	.09**	-.02	.01	-.01	-.02	.01	.-05**	
R_{AP}	-.09**	.25**	.03	.03	.04*	.00	.03	.01

**. Correlation is significant at the .01 level (2-tailed).

*. Correlation is significant at the .05 level (2-tailed).

Appendix F. Transcript of binary logistic regression on exclusion of references

	B	S.E.	Sig.	Exp(B)
Age of reference (relative to JA)	.00	.00	.34	1.00
Reference is a self-citation	-.45	.21	.03	.64
Reference is a journal self-citation	-.60	.26	.02	.54
Reference is un-published document	1.55	.19	.00	4.72
Citations of first author of reference	-.0001	.00	.00	1.00
Reference is published in an economics core journal	-.21	.10	.04	.81
The journal is an economics core journal	.31	.09	.00	1.36
Number of years from WP to JA	.17	.04	.00	.18
Reference is by author already cited in the document	.77	.11	.00	2.15
Reference is an unpublished self-citation	1.07	.47	.02	2.92
Constant	-2.47	0.14	.00	.08

Note: Number of observations: 3963

Appendix G. Transcript of binary logistic regression on inclusion of references

	B	S.E.	Sig.	Exp(B)
Age of reference (relative to JA)	-.04	.01	.00	.96
Reference is a self-citation	-.08	.17	.63	.92
Reference is a journal self-citation	.04	.20	.86	1.04
Reference is un-published document	..07	.28	.80	1.07
Citations of first author of reference	-.00005	.00	.09	1.00
Reference is published in an economics core journal	-.27	.10	.01	.76
The journal is an economics core journal	-.03	.09	.73	.97
Number of years from WP to JA	.31	.03	.00	1.36
Reference is by author already cited in the document	.73	.11	.00	2.07
Reference is an unpublished self-citation	.37	.64	.57	1.44
Constant	-2.29	.14	.00	.10

Note: Number of observations: 3968