THE SOCIETAL IMPACT OF OPEN ACCESS TO RESEARCH

A Dissertation

Submitted to the National Graduate Institute for Policy Studies (GRIPS)

in Partial Fulfillment of the Requirements for the Degree of

DOCTOR OF POLICY STUDIES

by

ElHassan Anas ElSabry

September, 2017

1.1 Abstract

Open Access to scholarly literature has been an issue under debate since the early days of the World Wide Web. For over two decades, study after study have investigated the impact of Open Access (OA) to research papers in a variety of different contexts. Those studies have predominantly focused on the impact it has on the academic community itself. For example, so many studies were published on what is called the "Open Access Citation Advantage (OACA)". They all tried to validate (or argue against) the claim that making an article openly available online results in more citations to that article, than its toll-access counterparts. Another large portion of studies have focused on business models associated with Open Access. Yet, implications of open access in non-academic contexts (practitioners, policymakers, patients, concerned citizens, etc.) have been the subject of many discussions and indeed was the basis for much of the advocacy work and its responses from science funding agencies, but rarely so in formal published studies. In fact, several researchers have specifically pointed to the lack of research in this area.

This study is the first comprehensive attempt to understand the potential impact of open access research on society at large (beyond academia). It starts by examining the rhetoric about societal benefits within the Open Access movement (as represented by advocates, policymakers and leading researchers). It goes on with collecting and synthesizing the scarce evidence available on the issue. The study then makes its contribution on the theoretical level by introducing a typology of the various science/society interfaces where access to research papers is needed. The proposed scheme is anticipated to provide guidance for future research

on the issue. In effort to add to this scarce evidence, the study also investigate more closely two specific groups of research users from outside academia. It seeks to understand factors affecting citations to open access journals in patents of US pharmaceutical companies and sheds light on the rising trend of "independent researchers" and how they might benefit from open access to scholarly literature.

Open access is an issue of growing policy interest. More and more governments and science policymaking bodies are involved in the Open Access debate. There is considerable pressure (from both sides of the debate) on policymakers to respond and take related decisions. The study aims to contribute a clearer picture of the non-academic uses of research papers and ends by recommendations to science policy makers as well as future researchers in the field.

1.2 Acknowledgements

I would like to start by thanking my mother who have always supported my education and encouraged me to continue to the highest levels of graduate study. She has sacrificed a lot to make sure I am raised to be a good man after the passing out of my father (may Allah bless his soul).

I would also like to acknowledge the support of my professors in the STI program. I would have not been able to achieve this without the kind support of Sumikura-sensei, my main advisor. He was very helpful at different stages of my work and was always cheerful and considerate. I also extend my thanks to other professors who have taken the time to advise me on different issues both related to my dissertations and outside of it. Thanks to Kuwahara-sensei for introducing me to the very interesting field of bibliometrics and for Patarapong-sensei for always emphasizing STI policy that is relevant to developing countries. I also want to thank Arimoto-sensei, Sunami-sensei and Suzuki-sensei for the fruitful discussions I had with them about different matters both related to this work and to other science policy topics.

I would like to also thank Professor Leon-Gonzales from GRIPS for kindly dedicating time to advise me on some parts of my work. I am thankful to other professors whom I have not (yet) met but where kind enough to provide some useful insights over email. These professors are John Willinsky from Stanford, Martin Eve from the University of London and Bo-Christer Björk from the Hanken School of Economics. Special thanks are due to Mike Taylor

from the University of Bristol for offering me space of his website (whoneedsaccess.org) to publicize an early version of the bibliography I collected.

Needless to say, thanks are also due to my dear wife Gehad who has been very supportive throughout our stay in Japan (very far from home) and to the little Salma for adding loads of joy to our life here.

1.3 Table of Contents

1.4 List of Figures

1.5 List of Tables

1 Background

1.1 Open Access Movement

The huge boom in the number of researchers and universities worldwide after World War II, partly because of the unprecedented levels of government funding for science and partly due to the need of establishing nationally-oriented universities in newly independent states, has called for an accelerated expansion in the scholarly communication enterprise [1]. This was mainly reflected in the takeover of the system by commercial publishers at the expense of traditional society publishers [2]. The introduction of financial interests to the system (as opposed to traditional motives like the pursuit of knowledge and intellectual merit) combined with the diffusion of neoliberal ideals in higher education in general has manifested in the continued rise of journal subscriptions prices, eventually leading to what became known as the "serials crisis of the 1990s". At that time the rate of price increase was so high (more than three time the Consumer Price Index according to some estimates [3]) that libraries around the world, especially in North America, started to cancel many of their subscriptions. In its report about the issue, the UK Office of Fair Trading noted that "there a number of features of this market that might militate against the operation of normal competitive market forces" [4]. It is no wonder that the top 100 publishers own about two thirds of the journals, while the top 10 alone own 45% of them [5]. Experts also estimate profit margins in the industry to be in the 20-30% range [6].

In response to this trend (and taking advantage of the concurrent expansion of the World Wide Web), a group or researchers and librarians started to coordinate their efforts in what

came to be called the "Open Access Movement". The Budapest Open Access Declaration in 2002 can be seen as the first major milestone of the movement. It provided both a definition of for Open Access and a roadmap for how the movement should go about to achieve its goals. Since that time, lots of discussions, declarations, workshops, conferences and lobbying have taken place in support of (or in opposition to) Open Access.

For research to be open access, two kinds of barriers have to be eliminated [7]. The first is the price barrier, which means that the research article should be accessible for any one with internet access at no extra cost. The second type is permission barriers, which are typically imposed by copyright (e.g. restrictions on printing, distribution and, more recently, text mining). Open Access can be achieved in two main ways, either by publishing in open access journals (*gold* route) or uploading copies of normally published papers in freely accessible archives (e.g. university repositories, PubMed Central, RePEc). There are a plenty of other "flavors" of Open Access [8], the most common of which is when a normally subscription journal offers its authors to make their individual articles open access in return for a fee (*hybrid* route). However, this hybrid model is very controversial. Many consider it "double-dipping" on the part of publishers, because in many cases the journal subscription prices do decrease even when many individual articles are freely available [9].

Although discussion of Open Access issues has been mainly among the communities of librarian's, scholarly publishers and to some extent academics, in the last few years more groups are getting involved in the conversation, not the least are science policy makers. The debatehas made its way to policy circles thanks to the "taxpayer right" argument [10]. Briefly,

it is the idea that for research that is funded by public money (the predominant source for funding for academic research), the results need to be openly available. Taxpayers should not be asked to pay again to see the fruits of their investment in science, let alone make use of it. This idea have fueled a separate line of argument. That is, who else "needs" access to research other than researchers in universities and research institutes?

1.2 Benefits of Open Access

For over two decades, study after study have investigated the impact of open access to research papers (as opposed to the traditional subscription system) in a variety of different contexts. The anticipated benefits of open access are widely viewed to fall under three categories [11]: benefits to the academic community, economic efficiency compared to the traditional subscription system and benefits to the wider society. The first two are briefly introduced below, while the third one is the main topic of this dissertation.

1.2.1 Academic Benefits

Many studies have predominantly focused on the impact Open Access has on the production and dissemination of knowledge among academics themselves. For example, many studies were published on what is called the "Open Access Citation Advantage (OACA)". They all tried to validate (or argue against) the claim that making an article openly available online results in more citations to that article, than to its toll-access counterparts. It is not strange that the number of studies about OACA has been increasing all this time given the extreme importance of citation counts for academic careers all along. Several bibliographies exist

trying to gather these studies and consolidate their findings [12,13]. The general understanding is that open access papers do in fact have a citation advantage over their counterparts behind paywalls. To some, this has already become an established fact [14]. Other than OACA, Open Access is thought to have many other positive impacts like the improved transparency in science, reduction of redundant research and making large scale text mining of scholarly literature possible [11].

1.2.2 Economic Efficiency

Another large portion of studies have focused on business models associated with Open Access. They mainly focus on the different economic implications of open access on libraries, publishers and science funding agencies [11]. Consensus is building around the fact that a fully open access system is generally more efficient than the current subscription tradition [15].There is, however, great variation in the proposed pathways by which this shift can take place [16,17].

1.2.3 Societal Benefits

Contrary to the two other types of benefits mentioned above, a small number of studies have been done to quantify or assess the impact of open access on the rest of society. In the open access movement rhetoric, the argument about freedom of knowledge is usually extended to mean that all society would benefit from the availability of research papers on the internet. Unsurprisingly, implications of open access in non-academic contexts (practitioners, policymakers, patients, concerned citizens, etc.) have been the subject of many discussions

and indeed was the basis for much of the advocacy work and many funding agencies' open access policies (as shown in Chapter 2), but rarely so in formal published studies. In fact, several researchers have specifically pointed to the lack of research in this area. They make statements like "little is known about the impact that free scholarly research literature might have on the knowledge and interests of laypeople" [18] or "almost no studies have evaluated whether free access to the scientific literature has had an impact […] in non-research contexts" [19]. A recent study aiming to describe the Open Access evidence base (i.e. research done *about* Open Access) has emphasized that the societal impact of open access "still needs to be systematically investigated and documented" [20]. A recent report has also asserted the existence of "a gap between the hypothetical societal good of open access and the minutiae of usage and interest measurements" [21]. The Research Information Network (RIN) report of 2014 has speculated the reason for this to be the inability to gather data on user demographics from currently available information sources (e.g. repositories and publisher platforms), which mainly focus on academic-related metrics (e.g. journal citations) [22]. Therefore, it appears that the issue of open access impact on society is not only an under-studied issue, but a difficult one to approach too.

1.3 Literature Review

This section is divided into two parts reviewing the necessary literature to provide the theoretical context for research questions asked in this dissertation. The first part reviews previous attempts to classify the different ways by which knowledge diffuses to society. The second part looks more closely into the issue of knowledge transfer barriers as inferred from

different available knowledge transfer models. This investigation is important to give context to the later contribution of this dissertation suggesting that the high cost of journal subscriptions might be one type of those barriers.

1.3.1 Diffusion of Academic Knowledge to Society

The question of the impact of open access on society (outside academia) is almost inseparable from the issue of how academic knowledge diffuses to society in general. This is why in this section a review of the models for general impact of research on society will be provided followed by more specific models that aimed at open access research in particular. Studies of the societal impact of research have mostly come in the context of justifying the increasing public funding for university research in the second half of the twentieth century. There has been a lot of efforts to measure this impact with the aim of informing policymakers on how to better utilize public funds. Bornmann's comprehensive review [23] provides a good summary of these efforts and the several challenges that lie ahead. However, the bulk of literature in the field of Science & Technology Policy (in some cases known as Innovation Studies) has sought to identify and measure the economic benefits of public R&D funding. In their seminal review, Martin and Tang [24] outlined seven channels through which the benefits of research can diffuse to the economy. Other than economic benefits, there has been a large gap in assessing different forms of societal impact of research partly because the problem of allocating public funds is inherently a political one [25] and partly because of the lack of proper indicators (and consequently data) to inform the debate. Meijer and colleagues [26] attempted to provide a more holistic view by classifying 4 types of returns scientific

research has. These are scientific returns (on the academic community), social returns (on professionals/practitioners), economic returns (through industrial production) and cultural returns (through laypeople's engagement with research). In another study [27] it was classified into societal outputs (e.g. products), societal use (references to research in practice) and societal benefits (i.e. changes in society as a result of research). Indeed, the introduction of UK's Research Excellence Framework in 2014 (with its emphasis on societal impact) has given a huge push to research on this topic, but much work remains to be done, at least compared to the much more developed "academic impact" evaluation tools [28].

In the context of open access, there has been three attempts to model the process through which the benefits of open access can flow to the wider society. The work of Zuccala [18,29] assumes that any discussion on societal benefits from open access has to happen based on what we already know from other fields operating at the science/society interface. These are information-seeking behavior, public understanding of science and science communication. In her model, Zuccala proposes that open access to research provides a more just and efficient way to communicate scientific knowledge to the general public: more just than the public education model (where self-selected intermediaries take on the mission of "simplifying science" to the public) and more efficient than the co-production model [30] which actively involves people in the scientific production process. In a completely open access system, laypeople have the freedom to directly engage with the scientific literature as part of their information-seeking activities. She maintains that it is still not clear the extent to which many people will appreciate this freedom or whether there will be a need for new mechanisms for mediation [29].

In his attempt to describe the scholarly communication system in terms of an IDEF0 process model [31], Björk five groups whose main interest is in applying the knowledge found in research papers (as opposed to producing it). These are universities (who use them to educate students), governments (by applying research findings in defining standards, granting patents and designing policies), companies (for product and process development), physicians (in treatment of patients) and private individuals (who use them to enhance their understanding of the world or their lifestyle. Houghton and colleagues further developed the model [32] and added NGOs (including lobby groups) as well as other groups of practitioners by giving the example of law and engineering professionals. They also added another type of usage for research papers. That is the production of secondary sources (e.g. blogs, textbooks and stories in popular media).

The most recent attempt to address the issue of open access societal implications was that by Bankier and Chatterji [21] in bepress's report. Based on the 100 stories selected for the report, they created a framework "to serve as a tool for stakeholders who are interested in advocating for open access on their campus yet lack the specific vocabulary and suitable examples". The highest level of the framework proposes three categories of benefits for open access: benefits to readers, authors and institutions. Of relevance here is the first category where 9 sub-categories for impact where presented. These are affecting public policy, advancing innovation, improving access to education, linking global experts, connecting cultures, building local community, informing patients & caregivers, updating practitioners and informing prospective (university) applicants.

The main issue with models addressing knowledge diffusion through open access is that they were not constructed based on the available evidence. They were developed using theoretical analysis only or some anecdotal evidence. The dissertation aims to close this gap by constructing the first typology of non-academic users of open access based on three different streams of evidence (more details in Chapter 4).

1.3.2 Barriers to Knowledge Transfer

Studies on knowledge transfer/diffusion have mainly targeted the issue of industrial application of academic knowledge, which is only one form of societal impact of research. Yet, it is the most extensively discussed one and the most relevant to the later part of this dissertation where the impact of open access on the pharmaceutical industry is studied in more detail (Section 5.1).

Several models have been developed to explain the process of knowledge transfer. These models were most recently reviewed by Abdul Wahab and colleagues [33] in 2009. Among these models, Eckl [34] suggests that two models can provide the basis for studying barriers to the knowledge transfer process. These are the stage model by Sung and Gibson [35] and the contingent effectiveness model by Bozeman [36]. In this first model, there is no much emphasis on the accessibility of academic literature to firms because the assumption is that "good ideas sell themselves" [37]. Regardless of anything, firms will successfully reach out to knowledge creators as long their product (the knowledge itself) is of good quality. This is a problematic assumption because it will not be easy for firms to identify which

researchers/universities are the most relevant to reach out to if all they can read is journal paper abstracts. This is especially the case in current days where the volume of published literature every year is huge (over 2.5 million in 2015 [38]) and efforts are already on going to produce algorithms to do literature reviews on behalf of humans [39]. Bozeman's model on the other hand explicitly mentions "open literature" as one example of knowledge transfer media (Figure 1-1). However, because it was seen as generally less effective than other (more direct) forms of interaction (e.g. contract research), it was not given enough treatment in the contingent effectiveness model discussions [36].

In attempting to express both linear and non-linear characteristics of the knowledge transfer process, the interactive-recursive model of knowledge transfer (Figure 1-2) suggests the this process happens as a temporal motion (linear part) in a three-dimensional space (non-linear part) [37]. The model was developed by Eckl in response to the shortcomings in Gibson and Bozemann's models as part of her study of knowledge transfer barriers. According to it, the three dimensions of the knowledge transfer process are knowledge creation, knowledge diffusion and knowledge absorption. All three dimensions interact with each other and the model specifies one determinant for the effectiveness of motion along each of the dimensions. For knowledge creation the determinant of success is the object of the created knowledge, more specifically, its quality and relevance. For knowledge diffusion, the determinant of success is how effective the medium is in making the created knowledge known to the circle of its potential users. Unlike the previous two determinants (where all participants have a role), the determinant for knowledge absorption (the absorptive capacity of the firm) is the mainly the responsibility of the knowledge taker [40].

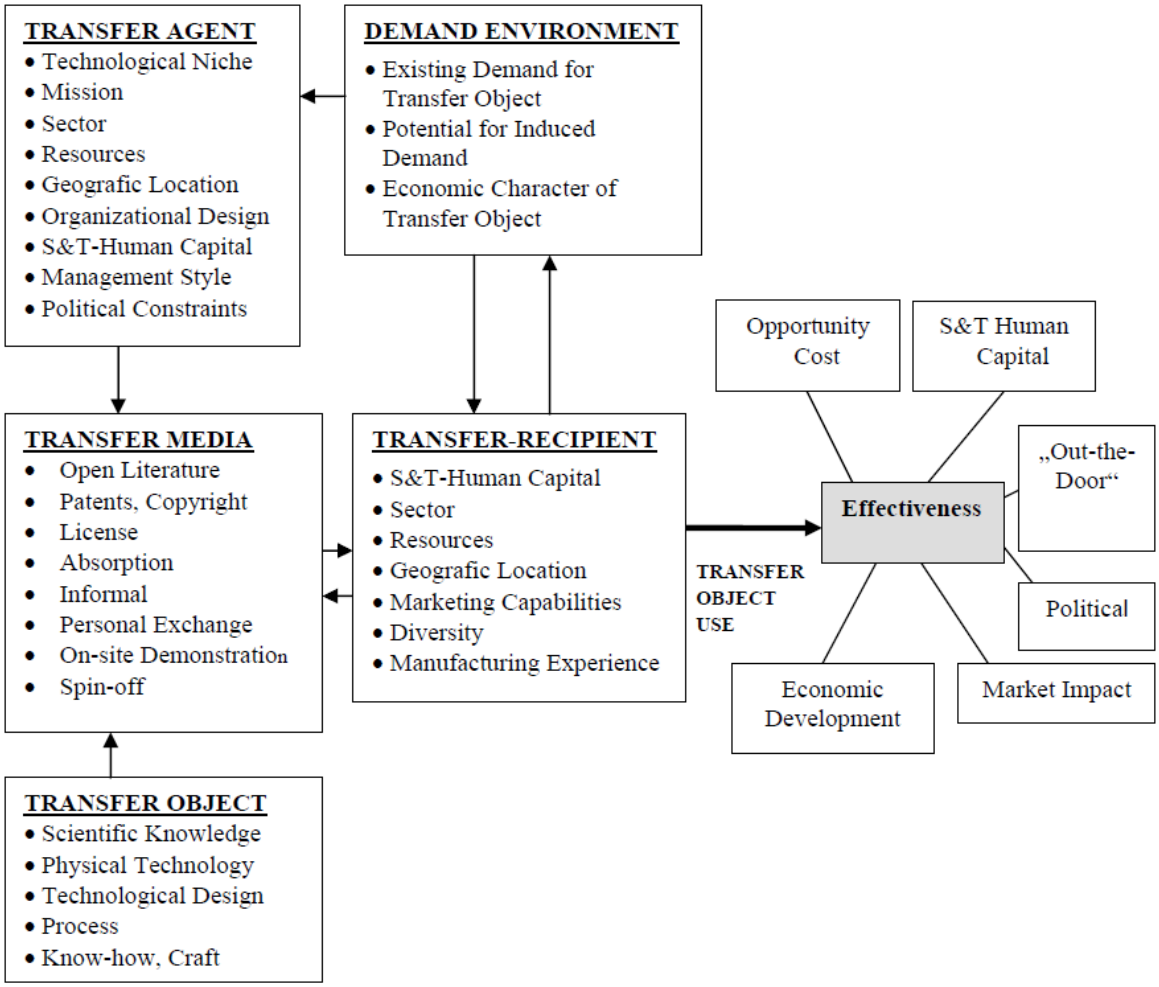


Figure 1-1: Contingent Effectiveness Model of Knowledge Transfer (source: Bozeman [36])

Barriers to the knowledge transfer process according to the interactive-recursive model take place as shortcomings in those determinants [37]. Of interest here is the issue of effective transfer media. As publication of research results is considered one of the main modes of knowledge diffusion, it is important to investigate the extent to which published academic literature is successful in reaching all potential knowledge users. With reports about an ongoing crises in access to literature in by SMEs [41,42,43,5], it is worth investigating whether high journal subscription costs can be considered one type of barriers to academia-

to-industry knowledge transfer. This is especially important in the light of the assertion by Cohen and colleagues that published research is among the key channels through which universities influence industrial R&D [44].

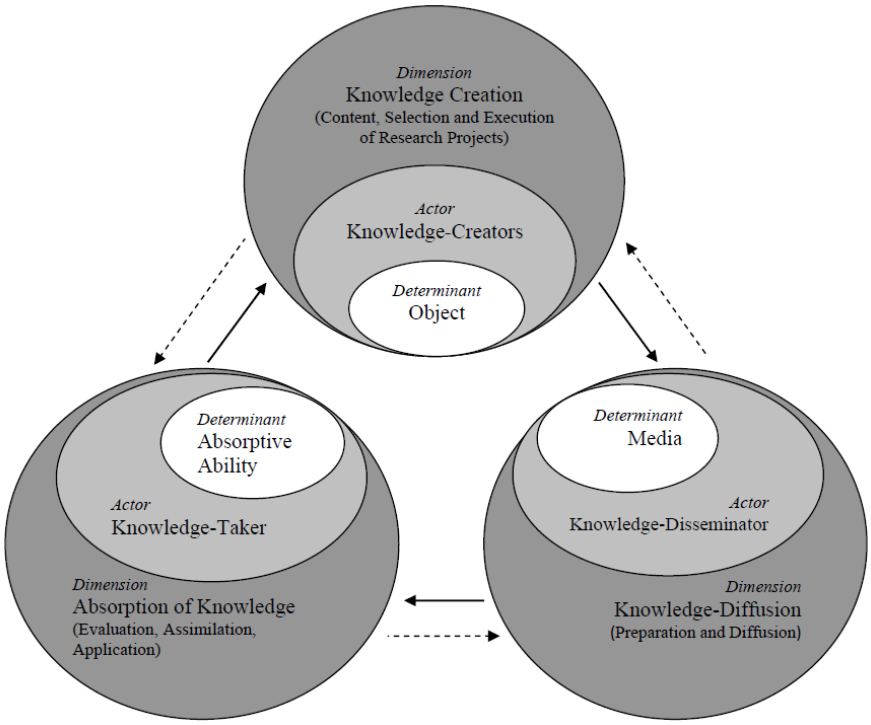


Figure 1-2: Interactive-recursive Model of knowledge transfer (source: Eckl [37])

Other attempts to identify barriers to knowledge transfer have considered financial constraints on companies as a possible barrier. Insufficient resources allocated to R&D was one of the seven barriers to knowledge transfer identified Irwin and More [45]. The bulk of the literature, however, is not on financial but on barriers related to organizational culture [46,47].

By investigating whether the usage of open access is related to the firm size, Section 5.1 in this dissertation proposes that the high costs of journal subscriptions might be considered a barrier to efficient knowledge transfer.

1.4 Research Purpose and Significance

The work described in this dissertation has two main purposes. First, it aims to explore the different uses of open access research in non-academic contexts and attempt to create the first evidence-supported typology of open access beneficiaries outside the academic community (as opposed to other less developed typologies reviewed in Section 1.3.1above). Second, it aims to enhance the evidence base available to support the typology. This is done through a deeper investigation of two groups of users from the typology. These are industry researchers and unaffiliated researchers. For industry researchers, the focus is on whether high journal subscriptions can be considered a barrier to knowledge transfer in the light of discussions presented in the literature review above (Section 1.3). For unaffiliated researchers, because little is known about them, this dissertation proposes the application of bibliometric techniques as a way to study these researchers.

It is crucial to note that having a good idea about the societal impact of open access (with strong supporting evidence) is not merely an issue of intellectual interest. Open access is an issue of growing policy interest. More and more governments and science policymaking entities are becoming involved into the Open Access debate and there is considerable pressures (from both sides of the debate) on policymakers to respond and take related

decisions. A common example of a related policy is the decision by a science funder to require grantees to make the papers emerging from their research openly available on the web. A well-known collection of those policies is maintained in the Registry of Open Access Repository Mandates and Policies (ROARMAP) database [48]. Figure 1-3 below shows the growth in the number of such policies as inferred from this database.

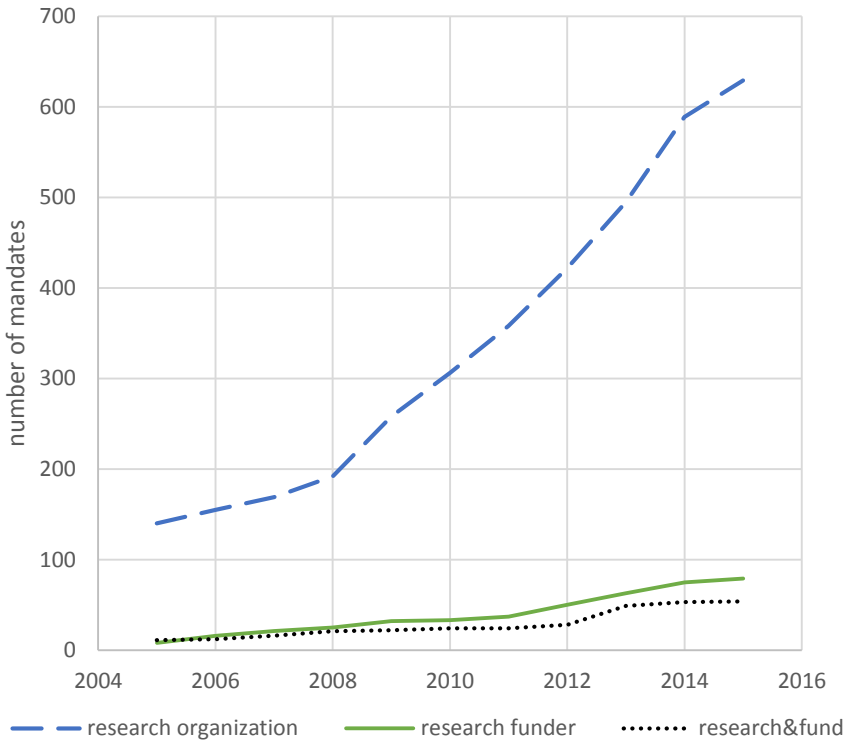


Figure 1-3: Growth of open access policies from different types of organizations according to ROARMAP

On a different level, Prosser has attributed this rising interest on the part of policymakers to a group of "political drivers", which he explained in two different studies [49,50]. To him, these factors exerted political pressure in support for Open Access. In the first study, he makes the argument that three drivers push the agenda for Open Access. These drivers are:

- the transition of several countries from industrialization to a knowledge-based economy;
- the rising emphasis on performance assessment to ensure accountability for governmental science budget;
- the global transition to a more online-collaborative research processes.

Prosser's second study (mainly based on policy developments in the EU) argued that the previous three drivers have intensified globally and added a fourth driver to the set. The additional political driver for Open Access was:

- the increasing social pressure for more openness in all public sector processes (e.g. Freedom of Information movement, Open Government Data)

Although it has been almost a decade since Prosser identified those drivers, they remain influential in open access debates nowadays. One could argue Prosser's point number 3 is not as relevant as it used to be (back in the days closer to the emergence and expansion of the World Wide Web). However, its essence is still to a large extent present as scientists keep demanding smoother, more efficient workflows from new technologies as they come up [51].

1.5 Research Questions

With those political drivers in play, along with the apparent lack of evidence to support the existence of societal benefits for Open Access, policymakers are in need of guidance to better understand this issue so they can make well-informed decisions. To achieve this, the

dissertation consists of two main parts. The first is an exploratory study that aims to examine the status quo and provides the groundwork for future research efforts. This exploratory study has one main research question (RQ). That is:

RQ1: Which groups of people outside the academic community benefit (or would potentially benefit) from open access to scholarly research?

In order to answer this question, two sub-questions are addressed.

RQ1a: Who is claimed to be a beneficiary of open access?

RQ1b: What evidence is available to support those claims?

The dissertation's second part empirically studies two groups of those identified in the first part, with the aim of contributing to the evidence base about the societal benefits of open access. The two studies address one of the following research questions respectively.

RQ2: How is the company size related to its usage of open access research?

RQ3: How can we study the usage of open access by unaffiliated researchers?

In response to these questions, the rest of this dissertation is structured as follows. Chapter 2 reports on a study that examined three types of documents issued in support of open access. The aim of examining these documents is to identify claims of societal benefits made by open access advocates, makers of open access policies and supportive leading researchers (RQ1a). Chapter 3 identifies and summarizes the supporting evidence about societal benefits of open

access pulled out after a bibliometric analysis of the landscape of research on open access in general (RQ1b). Chapter-4 merges findings from studies in Chapters 1 & 2 in addition to insights from previous literature of disciplines that investigate the science/society interface. The aim of this combination is to create a typology of all potential non-academic beneficiaries from open access, which would act as the general answer to RQ1, as well as a framework to guide future research on the issue of societal impact.

Chapter 5 focuses on two particular groups of users of open access research. Both groups were identified from the answer to RQ1. The first study (Section 5.1) uses statistical analysis of patent data as a way of understanding the how American pharmaceutical SMEs use open access research (RQ2). The second study (Section 5.2) sheds light on a new method that can be used to measure the impact of open access on researchers who are not affiliated with professional research organizations (RQ3).

The final chapter (Chapter 6) summarizes the contribution of this work and attempts to put it in context of science policy making.

2 Claims about the Benefits of Open Access

2.1 Introduction

This chapter presents the results of a study attempting to identify the different benefits of open access as anticipated by its supporters. Focus is mainly on claims about societal benefits of open access, i.e. those in non-academic contexts. As mentioned above, recent reviews of literature about open access [19,20] have emphasized the lack of enough research to

investigate the potential of open access to benefit individuals or groups of people who do not belong to universities and credentialed research institutes. This is in contrast to the abundance of studies about other aspects of open access (e.g. citation advantage). Davis & Walters [19] noted that "almost no studies have evaluated whether free access to the scientific literature has had an impact on the use of scientific information in non-research contexts such as teaching, medical practice, industry, and government". The reason for this has been speculated (by a recent Research Information Network (RIN) report [52]) to be that it is currently not possible to "gather systematic data on the demographics of users either on publisher platforms or via repositories".

This study comes in the context of a larger project [53] aiming to identify the societal benefits of open access and to devise new ways to measure and document this impact. Identifying the "claimed" societal benefits of open access (which is the aim of this study) will support the larger project in two ways. First, these claims can be tested against the currently available evidence about open access benefits to assess their credibility, which would be a useful exercise to guide the open access movement. Second, a deeper understanding of the discussion on societal benefits of open access will give insights about which stakeholders to include in the conversation and perhaps also inform the current debate on who should bear the cost of open access.

This chapter is organized as follows. An overview of the study design is presented at the beginning. Then, owing to the different types of documents analyzed in the study, the data used for each group of open access supporters is presented in a separate section along with a

brief background on the issue and the analysis results. A discussion of the overall outcomes and their implications then follows and the chapter ends with some suggestions for future research based on the study findings.

2.2 Study Design

A total of 164 documents were chosen to represent the views of a wide range of open access supporters. Focus was not on the entire content of these documents but only the one or more key statement(s) within them, where the purpose behind supporting open access was stated. While many of these "statements of purpose" were mentioned in a straightforward manner under a separate section of the document (e.g. "Why Open Access", "Advantages of Open Access"), some were spread all around the document and were inferred from the context. Three types of documents were analyzed for the purpose of this study. First, a selection of prominent statements and declarations about open access was used to represent the views of open access advocates worldwide. Second, policies in support of open access issued by government bodies were used to understand how policymakers perceive its potential benefits. Third, it was important to include the perspective of researchers. Hence, editorials announcing the launch of open access journals (or conversion of subscription ones to open access) were examined as a possible source for journal editors' beliefs about open access. While it can be argued that there is some overlap between these three groups of open access supporters, it is also important to note that the chosen documents were written for different purposes and address different audiences.

2.3 Claims about open access Benefits by Group

2.3.1 Open Access Advocates

It is very difficult to define “open access advocates” as a coherent group of people. It is a group that includes researchers, librarians, university administrators, research funders (both public and private), some scholarly publishers and even university students. Nonetheless, since what characterizes all of them is their outspoken support for open access, statements and declarations they produce can be a good representation of how they see open access and the benefits they expect from it. Declarations and statements in support of open access have played different roles at different points in the history of the open access movement. They were written to define the movement and lay out its main goals, to respond to related developments on the scholarly publishing scene or even to impose certain agendas on the debate. Many of them were used as tools to gather support for open access and were usually accompanied by large scale campaigns to call on people to sign them. While there is a wide range of documents fit the "statements in support if open access " description (e.g. one can consider every open access mandate or policy as such), a representative list of key documents had to be chosen for this study. The Open Access Directory (OAD) was consulted for this purpose. OAD is a community-sourced database aiming to document the open access movement. It is administered by a group of prominent open access advocates and hosted by the Simmons College.

Table 2-1: Beneficiaries of open access as mentioned in it key declarations

Statement	Year	Beneficiaries of OA
Tempe Principles for Emerging Systems of Scholarly Publishing	2000	researchers, industry, professors, students, informed citizens, the public
Budapest Open Access Initiative	2002	researchers, teachers, students, other curious minds
Bethesda Statement on Open Access	2003	researchers, developing countries, the public
Access to Scientific Information (by the Inter-Academy Panel)	2003	researchers, developing countries
Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities	2003	researchers, society
IFLA Statement on Open Access to Scholarly Literature and Research Documentation	2004	researchers, disadvantaged researchers
The Student Statement on The Right to Research (R2RC)	2009	researchers, students, patients, informed citizens, developing countries
Washington D.C. Principles for Free Access to Science	2004	not researchers , not patients

Eight declarations were selected from the OAD list of "Declarations in support of OA" [54]. They were selected owing to their significant influence on the open access movement, global nature and their representation of different stakeholders of the scholarly communication system (librarians, publishers, researchers, funders, students and prominent advocates). Table 2-1 lists the chosen statements, their respective years of adoption and keywords pointing to benefits of open access as declared by the statement authors/signatories.

As expected, all declarations assume that researchers are the main beneficiaries from open access. These benefits take two forms, either through direct gain (i.e. visibility and citations for one's own work) or as general enhancement of the quality of research through the

transparency and democratization offered by open access. Four declarations have made reference to the subgroup of "disadvantaged researchers". Those are researchers whose institutions could not (or have never been able to) cope up with the rising costs of access to journals, especially in developing countries. The Budapest Declaration later specifically argued that open access should not be understood as a one-way communication tool (i.e. from the knowledge-rich north to the knowledge-poor south) but as providing mutual benefits for both. The declaration argues that removing barriers to scholarly literature will "share the learning of the rich with the poor and the poor with the rich".

References to the benefits of open access to the educational process was made in three of the declarations. Needless to say, students' frustration with the lack of proper access to research was the main driver for the R2RC Statement. One of the signatories (The European Federation of Psychology Students' Associations) has even tried to systematically study this a few years later [55].

Beyond the academic/research community, three declarations made reference to "society" and "the public". It is not clear if that was intended to mean specific benefits to laypersons from access to scholarly literature, or indirect benefits from an enhanced body of knowledge. One of these declarations, the Tempe Principles, also mentioned "an informed citizenry and a healthy global economy". However, these were mentioned as the outcomes of "creation, dissemination, and application of new knowledge". In other words, they are not direct beneficiaries from access to literature. One could also argue that Budapest Declaration's reference to "other curious minds" is intended to encompass any groups of potential users

outside academia. Nonetheless, none of the sources later cited in the declaration (as a proof of efficiency of open access) provide information about usage outside academia.

The R2RC statement was the only one to make explicit reference to some groups of society that could benefit from access to research. They mentioned patients who would "have access to the latest medical research" and citizens who could "evaluate scientific information on environmental impacts". The statement does not offer any supportive evidence in this regard. However, as the most recent declaration among the ones in this study, it is possible that some results of research about of the open access impact on society was already available to those who drafted it. This possibility is strengthen by the fact that four of the six open access policies cited by the R2R statement make similar claims about groups of societies that could benefit from access to scholarly literature. These include clinicians, policymakers (CIHR policy), families, patients (Autism Speaks policy), media (Canadian Cancer Society policy) and educators (Stanford GSE motion).

While many would consider the "Washington D.C. Principles for Free Access to Science" not a statement "in support" of open access as defined by this study, it was nonetheless important to include it here, mainly because this particular declaration argued against the societal benefits of open access. In addition to denying the need for access (even among researchers) by claiming that "published literature is routinely and readily available to all who need and want it", the declaration asserted that "[it] is debatable whether members of the general public can actually benefit from reading the original research literature, as its arcane and specialized reporting is intended primarily for other researchers,". This was also

extended to imply that benefits of open access to clinicians is also debatable given that "many findings are not relevant for immediate clinical application". A few years later, some still maintained that there is no evidence for "unmet demand for the primary medical or health sciences literature among the general public", albeit this "does not necessarily reflect the absence of unmet demand" [19].

Examining these key declarations shows that the issue of societal impact of open access was not strong on the agenda of most open access advocates who drafted them. Their main contention appeared to have been that it was not wise to ignore the value internet can add to scholarly communication. Open access was the most efficient way to ensure the freedom of knowledge and internet's contribution to enhancing research quality and reach.

2.3.2 Policymakers

There have been several case studies published to examine the impact of open access policies and mandates on individual institutions. They examined things like policy efficiency [56] or the researchers' reaction to these policies [57]. Only a small amount of studies aimed at studying *government* open access policies though. PASTEUR4OA (Open Access Policy Alignment Strategies for European Union Research) is probably the most extensive research project concerned with open access policies. It is based on the same database of open access policies used in this study and analyzes different policies. The aim was to encourage EU member states to align their open access policies to ensure best practice and to make

compliance easier for researchers funded by grants from multiple source. Addressing policymakers about the anticipated benefits of open access, PASTEUR4OA researchers referred in several documents (for example [58]), to open access as a strategy to cut publishing costs and a way to foster innovation by giving SMEs access to the scholarly literature. In fact, one of their policy briefs was dedicated to present a framework of how knowledge transfer (via open access) has "spillover" effects on many segments of society outside the research community [59]. Policy guidelines developed by UNESCO have also echoed similar arguments but also emphasized the impact of access to biomedical literature on patients and healthcare practitioners [60]. In this regard, Waltham noted that in the US this tendency (to encourage public access to research) is a result of pressure by patient advocacy groups, in the UK it stems from a more general mission to raise the public understanding of science [61].

The Registry of Open Access Repository Mandates and Policies (ROARMAP) is a well-known, comprehensive resource for open access policies from organizations all over the world. However, the majority of listed policies are issued by universities or research units requesting (or requiring) their faculty and research staff to make the outcomes of their research openly available on the internet. As of December 24, 2016, only 136 (of around 800 listed polices) were issued by organizations described as funders (82) or organizations the both perform and fund research (54). These were either government bodies (e.g. ministries, parliaments), national research councils, national academies, or other smaller units. Twenty-nine private research funders (e.g. Wellcome Trust) were excluded as well as four entries that were not government bodies but partnership programs or universities.

Of the remaining 103 public research-funding organizations listed in ROARMAP, some fell under policies of larger organizations (16), issued policies that were not about research papers (9, e.g. open data polices), or published other types of documents (4, e.g. not a policy but workshop recommendations). Three policies could also not be found. Therefore, analysis for this study was based on 72 policies that fit the initial criteria. The majority of policies had some English version available online. For those that did not (12 policies), online automatic translation was used to identify and translate the statement of purpose in the policy. A native speaker was consulted in cases where the automatic translation was not clear. Table 2-2 presents the overall results of analyzing the 72 valid policy documents.

The majority of policies (61%) make at least one claim about the positive impact of open access on the research community. This does not seem to be different whether the organization issuing the policy is only a funder or also conducts in-house research. This is understandable given that benefits to the research enterprise can safely be considered the main purpose behind all of these policies. As mentioned before, much evidence has been piling up over the years to support the belief in open access benefits to the research enterprise. Examples of this positive impact include enhancing the quality of the research, allowing for more reproducibility, avoiding duplication of efforts and supporting the globalization of science with more reading and citations.

Table 2-2: Beneficiaries of Open Access according to government policies

Benefits of OA	**Sample Keywords**	**Frequ-ency**	**Percen-tage**
research	quality, impact, reproducibility, duplication of efforts, open science, globalization, pace	44	61
industry	economy, growth, (open) innovation, valorization	34	47
public	awareness, culture, public understanding of science, taxpayer right, public accountability, scrutiny	29	40
professionals	users, deployment of research, uptake, clinicians	10	14
government	policymakers, public sector	10	14
education	OER, educators, fast percolation to high education	9	13
credibility	evaluating program managers, government transparency, M&E, efficient use of funds	8	11
visibility	intellectual gap, global recognition	6	8
NGOs	charities, NPOs	4	6
no mention		17	24
Total Number of Policies		72	100

An interesting finding from analyzing these government open access policies is their consistent emphasis on the benefits of open access to the economy. This is not about Open Access being economically more efficient by some "system-wide cost savings", as it was shown by Houghton [62] for example. Rather, it is about open access making more knowledge available to firms to build on, creating innovative products and services that

would consequently boost the economy. The argument is well summarized in the European Commission's position that "[fuller] and wider access to scientific publications and data … help to accelerate innovation" because "faster to market = faster growth" [63]. A similar sentiment can also be detected in the US government commitment to fund and make available research which "catalyzes innovative breakthroughs that drive [the American] economy" [64]. This is also consistent with Prosser's idea that the move to more knowledge-based economies is one of main drivers supporting the argument for open access among policymakers. He mentions that "[as] developed countries struggle with the transition to post-industrial economies, there is a growing belief that knowledge provides both power and economic growth" [49].

Policymakers concern about benefits to the taxpaying public is understandable. Of the 40% of policies that mentioned these benefits, some made broad claims like preserving knowledge and culture. Open access would enhance the knowledge produced by researchers and allow for maintaining it, which consequently will make it more relevant and useful to society as a whole. What lacked evidence was the more specific claims made by other policies about open access making possible the public scrutiny of the research outcomes. It is not clear what mechanism this will happen through. Indeed, some policies mentioned that open access will allow for better evaluation of the funding programs and their managers, which will consequently result in more credibility for the organization. However, this kind of benefit was included separately under "credibility" because it is more about accountability to other (superior) bodies of government than to the public *per se*. The claim that open access will increase the public understanding of science is also one that lacks supporting evidence. Even

participating in citizen science projects does not guarantee an increase laypeople's understanding of science [65], let alone the mere presence of scholarly literature online.

Three other benefits get nearly equal attention from policymakers. These are open access research usage by practitioners (e.g. doctors, lawyers, etc.), usage by public sector researchers (e.g. policy research units) and the open access as a form of Open Educational Resources (OER). Each one of these benefits is acknowledged in one sixth of the policies.

The relatively high percentage of policies (24%) that mentioned no specific purpose for supporting open access is mainly because some policies were not issued in a separate policy document (e.g. law, resolution) but as a changes to already existing documents that included more topics than just open access (e.g. national science law, guidelines for using research funds, etc.).

2.3.3 Leading Researchers

As mentioned above, when considering the "anticipated" benefits of open access, the opinion of researchers cannot be dismissed. This is especially true for researchers who have leading positions in their fields. Editorials are by definition a good venue where journal editors can express their thoughts of ideas about different issues related to their field. For this study, a set of 85 editorials were collected (where a new open access journal is announced or when an existing journal announces conversion to open access) to determine the views of leading researchers about the benefits on open access. These editorials come from journals across different fields of research (albeit with very strong presence of biomedicine).

The selection was based on searching the content of Scopus database as of January 21, 2016. Search was limited to publications of the type "editorial", which contain the expression "open access" either in the title or in the indexing (or author-provided) keywords. After excluding publications where "open access" was used to describe an unrelated concept (e.g. open-access endoscopy or open access railway infrastructure), a list of 517 editorials remained (including 15 duplicate entries). Titles of these editorials were then checked to classify the editorials into four groups:

- editorials announcing a new open access journal or a subscription journal's transition to open access (85)
- editorials announcing some new green or hybrid open access policy (60)
- editorials discussing open access without announcing OA-related decisions (257)
- editorials whose topic is unidentifiable based on title (100)

Only the first of these three groups was used in this study as a source of journal editors' views on open access benefits. The second group was excluded because only reading the editorial might not have been enough to know the real intentions of choosing open access. It is not clear too if the purpose was anything more than compliance with funder requirements or the increased revenue associated with the hybrid model. The results for analyzing the first group of editorials are summarized in Table 2-3.

Table 2-3: Beneficiaries of open access according to journal editors

Benefits of OA	Sample Keywords	Frequency	Percentage
wider dissemination	exposure, impact, visibility, indexing, archiving, citations, author retains copyright,	62	73
efficiency	access for developing countries, disadvantaged researchers, freedom of knowledge	33	39
rapid publication	immediacy, competitiveness	26	31
professionals	practitioners, clinicians, stakeholders	18	21
rising trend	citing open access declarations, compliance with funder mandates, revolution of scholarly publishing	17	20
public	taxpayer right, public understanding of science, interested laypersons	15	18
government	evidence-based policymaking	7	8
other groups	amateurs, media, parents, teachers	3	4
industry	drive innovation, private sector R&D	4	5
education	university students, professors	2	2
no mention		10	12
Total Number of Editorials		85	100

The most significant result from analyzing the open access benefits as seen by journal editors is their consistent focus on benefits to the research community. Unlike the previous two types of documents, editorials refer very little to any "public" or societal benefit of open access (only 18%). Even groups of people who might not necessarily be part of the research

community but are very close to it (e.g. practitioners 21% or students 2%) are mentioned relatively very little. Benefits to industry are also rarely mentioned, although most of the editorials come from the field of biomedicine, which is traditionally associated with the very "science-intensive" pharmaceutical industry.

Another interesting aspect is that only six editorials (7%) made reference to compliance with funder mandates. This suggests that (at least for gold open access journals) a move to open access in communicating research might have happened naturally even in absence of funder mandates, given that editors chose to emphasis other benefits of open access.

Otherwise, the great emphasis that the majority of editors put on benefits to researchers in their field as the primary reason for support open access is very plausible. This was especially true for new journals that tried to emphasize benefits like citations and exposure as a way to attract their initial submissions (sometimes in combination with other strategies like waived APCs). It is however important to consider that for some journals the move to open access was also the move to online publishing, which by itself can account for benefits like rapid dissemination or more global visibility (relative to print-only publishing).

Some editorials mentioned adopting open access would be “sponsored” by a parent organization, i.e. no APCs will be required. However, it remains a limitation of this study the inability to know if the perceived financial gain from APCs was the main reason behind choosing open access. This is especially important to consider in cases where the editorial mentioned open access as a way to support the “growth” of their journal.

2.4 Discussion

Comparing the position of each of the three groups of open access supporters signals two main differences. One difference is that ideas about what benefits open access has on the researching community seems to be more about the researchers themselves (e.g. citations, visibility, copyright ownership, etc.) as viewed by journal editors, while at the level of policymakers more "abstract" benefits are generally perceived (e.g. globalization of science, reproducibility, transparency, etc.). Declarations occupy a somewhat middle ground on this issue. Regarding the benefits to developing countries, policymakers are the least to refer to this point. However, five policies (coming from Ireland, France, Brazil, Belgium and Slovenia) make reference to the somewhat similar concept of bridging the intellectual gap by making their own research more visible.

The analysis has also shown that there is near consensus that benefits of open access go beyond the academic/research community. Still, there is a lot of variation among the three groups in how they perceive the extent and reach of these societal benefits. The little regard open access journal editors give to open access benefits beyond the research can be explained in two ways. It is possible that they do not believe those benefits exist. This is understandable given the very little research done on this issue. Supposedly, researchers are more inclined to make evidence-based claims than most activists and policymakers. The other possibility is that they believe those benefits exist but (in writing those editorials) chose to focus on benefits to researchers as a way to garner support for their decision to adopt open access. In

both cases, more research is needed on this topic to inform researchers about any potential societal benefits for open access, which in turn might influence their decision to adopt it.

2.5 Conclusion

Investigating the claims about societal benefits of open access in different types of documents contributes two main observations. First, there is an abundance of claims about groups of people who would benefit from the large-scale provision of open access research. According to the investigated documents these groups are:

- industry researchers
- practitioners (mainly in medicine but others like teachers were mentioned too)
- policymakers,
- researchers in nonprofit organizations
- individual laypeople (both as patients and as citizens who want to become well-informed about different issues)

The second observation is that none of the documents states any evidence to support these claims. This is especially important because it is true that some documents refer to evidence supporting claims about other types of benefits (e.g. citation advantage), but not societal benefits. In this regard, research is needed to identify, classify and compare any literature that investigated the impact of open access on society. This should be done in order to (1) see if there is enough evidence to support claims discovered by this study and (2) identify if there are more groups of society that were shown to benefit from open access research. This

search for evidence is the main topic of the following chapter (Chapter 3) and groups of beneficiaries of open access identified by both studies (from Chapter 2 and Chapter 3) are then combined to develop a typology of open access beneficiaries to provide guidance for further research (Chapter 4).

3 Evidence Base for Societal Benefits of Open Access

The main purpose of this chapter is to explain the process of collecting and synthesizing the available evidence on the societal impact of open access. It starts with an analysis of literature about open access in general (Section 3.1) then proceeds with a deeper investigation of the content and characteristics of literature about societal impact in particular (Section 3.2). The literature analyzed in the second analysis is partly extracted from that found in the first one in addition to other sources.

3.1 Landscape of Research *about* Open Access

The Open Access movement has an interesting tradition of creating lists. There are lists for important dates in the open access movement and for the blogs that discuss open access and many other things. There is a (black) list of publishers of open access journals that do not align with certain quality criteria. The Open Access Directory (OAD is the resource to go to for exploring the vast majority of those lists [54].

Of interest to this study is the collection of lists compiled by Charles Bailey on scholarly communication and open access. More specifically, his two lists published in 2005 and 2010 with the titles “Open Access Bibliography” [66] and “Transforming Scholarly Publishing through Open Access” [67] are the very famous bibliographies of literature on open access. They both try to gather all research papers, commentary, and discussions about open access from a variety of different scholarly sources. Other Bibliographies also exist to identify research papers about specific aspects of open access. These include many by Bailey himself

but there are other notable examples. Indeed, one of the most influential bibliographies is "The effect of open access and downloads ('hits') on citation impact: a bibliography of studies" curated by Steve Hitchcock [13]. These bibliographies were put together based on the assessment of their authors who search different relevant venues for developments in the field of open access. Another type of collections of research about open access comes in the context of bibliometric studies that create the list of articles by searching one database using keywords relevant to open access. The databases they use are Scopus [20,68], Web of Science [69], and Library and Information Science Abstracts [70].

This section summarizes the results of a study that aimed to understand the characteristics of research published *about* open access. The papers were collected using the Scopus database but manual verification for each record was necessary to ensure the following:

- publications from different disciplines (not just library and information science) are included but without irrelevant papers (e.g. open access alimentary endoscopy)
- documents are correctly classified by type and not merely following Scopus classification (which was shown to have some errors)
- a paper that does not report the findings of a study is not included in the Original Research category (even if it was categorized as a research paper before)

There is a need for good understanding for this landscape of research to ensure that related policymaking is evidence-based. For the purpose of this dissertation, there is the added value

of making the task of identifying the pool of papers of interest (those studying societal impact) much easier.

3.1.1 Methodology

A simple search in the Scopus database was performed for the expression "open access" (along with its other variation, open-access). Restriction was applied to include only articles where the expression appears in the title or among the publication's keywords. The latter includes both author-provided keywords and those assigned to the publication on indexing in Scopus. There was no restriction on any other field (e.g. year, country…etc.). The decision to search with titles and keywords only (i.e. not to search within the abstracts, which are also indexed in Scopus) was to ensure that the number of results is reasonable to handle by manual verification in later stage of the analysis. It was also believed that most of the significant literature about open access would include the expression in their keywords if not their titles (in case fancy titles were used).

3.1.2 Results and Discussion

A total of 5,206 documents were retrieved from the search (over 33,000 in case searching within abstracts too) on January 21, 2016. Table 3-1 describes the steps towards refining this set of documents to reach the different sets of documents used for analysis in this study.

Table 3-1: Summary of the dataset at different stages of analysis

Action	Excluded Documents	Remaining Documents
searching for documents with "open access" in their title or keywords	n/a	5206
excluding documents with "open access" not in the context of scholarly communication (e.g. open-access liquid chromatography, Open-access endoscopy)	2,129	3076
restriction to documents in English	307	2769
restriction of documents published in journals (excluding books, chapters and conference papers)	476	2293
identifying original research papers	1235	1058

A number of important remarks should be made about the refining process. First, documents with the term "open access" used with a different meaning accounted for over 40 percent of the initial results. This is important to take into consideration for researchers who conduct bibliometric analyses based on specific keyword search. This subset also includes a group of 11 documents published between 1898 and 1906 referring to "open-access libraries" as well as several entries on open access in higher education (e.g. "Navigating 'open access' community colleges: Matriculation policies…."). Researchers have to be cautious even if the article's research field (which is usually determined one the journal level) close to the topic they are interested in.

Second, the low number of non-English papers should not be assumed to reflect the reality about writings on open access. It is often the case that authors would use keywords in non-English languages if the full text is also non-English. Indeed, regions like Latin America (which is extremely active in the Open Access movement) would have a lot more contributions that those represented here as they tend to use other keywords to describe it (e.g. “acceso abierto”).

Third, the analysis was restricted to journal articles because (in most cases) publications like books are already based on synthesizing findings from journal articles. At the same time, conference papers in many fields usually report preliminary findings of research that are consequently published as journal articles. Figure 3-1 shows the breakdown of publication types in the original dataset (after excluding irrelevant documents).

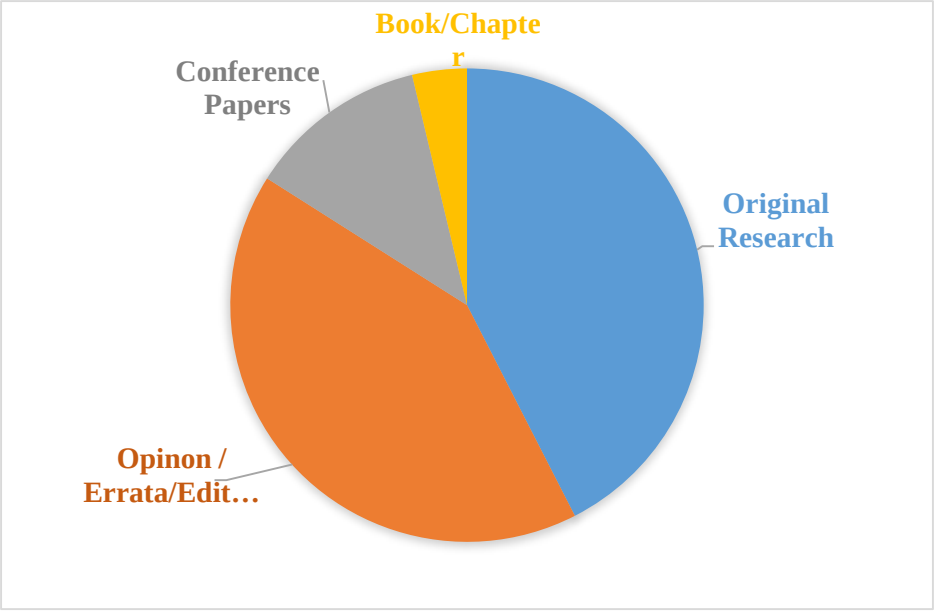


Figure 3-1: Types of documents extracted from Scopus with "open access" in their title or keywords

Among journal articles, distinction was made between original research papers and other types of articles (e.g. editorials, opinion pieces, etc.). It was based on manual revision of the contents of abstracts of the 2,293 journal publications, while consulting the main text if the abstract was not clear. This was a tedious, yet essential, task given that the main purpose of this study was to single out available substantial evidence about open access.

In Figure 3-2 the proportions of original research and general discussions are shown for journals of each field of research where there is active conversation about open access. It can be observed that for some fields the volume of discussions about open access far exceeds research evidence. This is partly because of the general trend of unsubstantiated claims within the debate on open access (Section 1.1) and partly because the debate is relatively new to many fields of research, which would naturally call for some debate before any serious research is undertaken. The prevalence of discussions in the biomedical field is very clear. This makes sense given that the debate is most severe within this field given the unreasonably high prices of some of its journals. It does however call for more attention to be given to creating more evidence to debate around instead of mere personal opinions. As expected, most research is done in library and information science journals.

As mentioned previously stated, further search for evidence on societal impact was based on extracting papers from the list of 1058 original research papers. In what follows (Section 3.2), more details is provided.

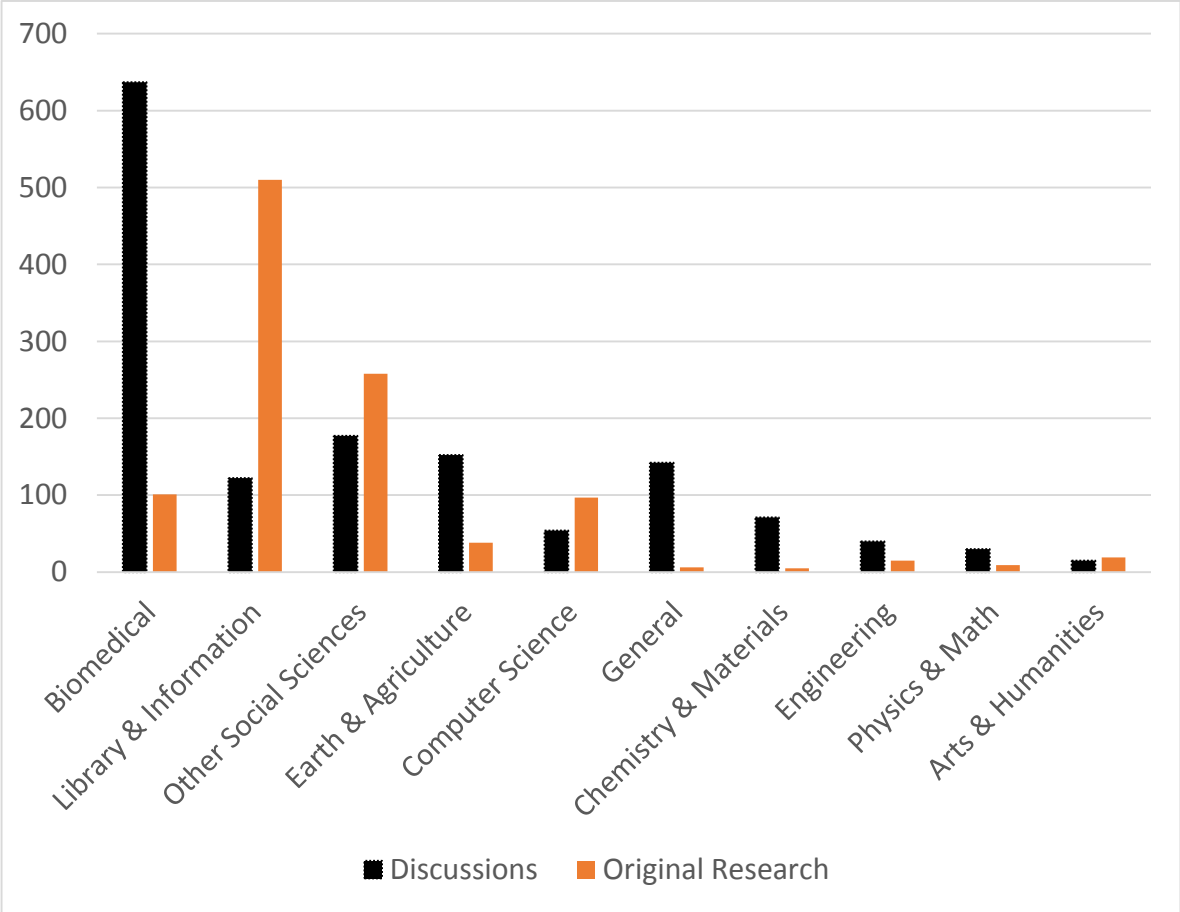


Figure 3-2: Documents discussing open access versus actual research papers in different research fields

3.2 Societal Impact Evidence Base

A total of 52 studies were found to address the issue of open access societal impact. The collection process was based on three strategies:

- eyeball search within the set of journal papers (both original research and discussion documents) from the previous study
- relevant keyword search in Scopus database and Google Scholar.
- citation tracking (both forward and backward)

Table 3-2 provides a description of the collected literature. It is noticeable that most of it is relatively recent. Government commissioned reports are the second most common type of documents making them a very important source of information about different societal impacts. It is worthy to note the prevalence of documents that do not describe substantial research projects and only provide some anecdotal evidence.

3.2.1 Open Access beyond Academia

There are two types of studies dealing with the societal impact of open access: studies that focus on one particular group (outside academia) and how it can benefit from open access and studies that seek do not target any specific group, but approach people outside academia in general and study their relationship to open access. This section will focus on studies of the latter type, followed by a separate section for each of the groups from the first type.

The two most detailed attempts to assess the wider impact of open access were Alperin's dissertation on the "The Public Impact of Latin America's Approach to Open Access" [71] and the "Citizens Demand for open access to Academic Papers" by Sato and colleagues [72] in Japan. As obvious from the title, Alperin's dissertation was mainly focused on Latin America and used its two most prominent open access portals, SciELO and RedALyC. The idea was to collect article-level metrics from these portals (along with other altmetric sources) and combine this with demographic data of the person who reads the article.

Table 3-2: Breakdown of studies on societal benefits of open access

Property	Category	Count
Publication Year	2001 - 2005	5
	2006 - 2010	11
	2011 - 2015	29
	2016 - …	8
Document Type	journal / conference paper	30
	dissertation / book chapter	3
	report	12
	opinion article (in academic journal)	8
Methodology	multiple methodologies	8
	theoretical analysis (including review studies)	8
	survey	10
	interviews / focus group	3
	bibliometric analysis	8
	experiment	2
	case study	3
	anecdotal evidence	11

Demographic data was collected using pop-up surveys at the time of download in addition to a small scale survey of those who shared Latin American research articles via Twitter. Results show that students (40% undergraduate, 60% graduate) are the primary user group of Latin American research, accounting for 44.5% of survey respondents. This is followed by university employees (including faculty) at 20%. Users who are unaffiliated with university (i.e. from outside academia) make up the rest of the survey sample at about 35%. Roughly 40% of these were public sector employees, 40% from the private sector and 20% from

nonprofits. In a different context, with the aim of identifying their perceptions and experiences with open access, Sato and colleagues [72] surveyed 800 Japanese adults. Respondents were a balanced group of people with higher education degrees (50%) and those without (50%, including university students). The sample included slightly more males than females with more people from the 30-49 age group (50%) than from other younger (25%) or older (25%) groups. The majority of respondents (55%) claimed that open access is useful or slightly useful to them. The top two reasons they give for their need for access are "satisfying curiosity" and "research articles being a credible source of information". Those who did not find open access useful ranked "the gap between academic research and daily life" and "the difficulty to comprehend academic jargon" as their top reasons.

Both studies document the persistent interest in health-related research by non-academic users. This is in agreement with results from a smaller scale (focus group) study of 23 Dutch citizens reported by Zuccala [18]. Participants argued that access to medical research can be of more value to lay readers as it relates more to people's lives (more than other types research e.g. mathematics). Echoing the Japanese survey participants, this Dutch group voiced concerns about the layperson's ability to comprehend academic jargon and showed preference for "human sources of information", but still appreciated research papers as a credible source of information within the flood of other sources on the web. A different group of lay people (in the United States, as surveyed by Harris Interactive in back in 2006 [73]) offer a different line of reasoning by claiming that, by virtue of being mostly tax-payer funded, medical research has to be freely available online to doctors and to those with special medical conditions. Over 80% of 2,501 respondents made this claim. Also in the United States (in a

Pew survey [74]), 26% of those who search for online health information claimed to have hit some sort of a paywall. Only 2% of those faced by the paywall decided to actually pay.

Other than medicine, it is not clear what other disciplines could be considered "people-related". Nonetheless, the assumption (by Smith [75]) that an open access political science journal will enhance public engagement with the discipline, indeed deserves consideration. Willinsky has also made the case for the importance of open access to philosophical literature based on Derrida's notion of the right to philosophy [76].

Using access logs of the Kyoto University website, another study [77] in Japan aimed to identify external links that refer to papers deposited in the university repository. It was discovered that, although not huge in number, a remarkable variety of websites linked to these open access papers including blogs about personal hobbies, websites by patients or their families, Q&A website and Wikipedia. The impact of open access to research on Wikipedia (as an intermediary supporting research diffusion to society) has itself been the main subject of two studies. In the first one [78], Willinsky randomly selected 100 Wikipedia entries to analyze the external references cited in them. The objective was to identify how many of these references were open access and later (for a subset of 20 entries) whether other open access references could be found online to cite for each entry. The second study (by Teplitskiy and colleagues [79]) offers a more detailed analysis of the relationship of paper's accessibility and being cited in Wikipedia. By the statistical analysis of data matched from the Scopus database and citations in the English Wikipedia, they find that a journal being open access increased it odds of being cited by Wikipedia by 47%. It is worth noting however

that their choice of journals as the their units of analysis (as opposed to individual articles) underestimates the impact open access can have on diffusing science through Wikipedia because it neglects articles that are openly available through the green or the hybrid routes.

Other researchers, attempting to cover the whole range of impact open access has, provided a more superfluous treatment of societal impact in particular. Tennant and colleagues [11] provided several examples of societal groups that could benefit from open access (including patients, NPOs, businesses and citizen scientists). On the other hand, the "100 Stories of Impact" report by *bepress* showcased 34 stories (collected from university repositories using their Digital Commons software) where access to research had significant impact on non-academic readers, although 19 of these were not about open access to than research papers (e.g. a database of tractor testing reports or local community volunteering to digitize a historic newspaper) [21].

3.2.2 Industry Researchers

Companies, large or small, run the most active research facilities outside the traditional centers of research (i.e. universities and research institutes), sometimes even with much larger budgets. This probably entails a huge need for access to research. Accessing literature in the form of journal subscription might not be ideal for companies, given that they frequently change interests according to technological and market changes in their business [serials crises]. Still, 15-17% of publisher revenues come from corporate subscriptions (STM, 2015). Indeed, different companies have different information needs. They also have

different levels of absorptive capacity [40] in dealing with knowledge sources from outside the company. Such differentiation results from the variation in sector, size and management style. The variation in sectors is probably the most influential, especially with the rise of the so-called "knowledge-intensive" industries in the past few decades. According to Prosser the movement towards the knowledge economy have been one of the main drivers behind open access policies adopted by different governments [49]. Indeed, this has been evident in the language many of them use in their open access policy documents (see Chapter 2). Picarra explained this impact on open access on the economy in terms of "spillover effects in all sectors of society" which drive "economic, social and technological progress" and laid down different arguments on why it is very essential for Europe's SMEs [59]. Small and Medium Enterprises (SMEs) in particular are thought to be at a disadvantage given the huge expenses they need to spend if they want to ensure access to all the literature they need [41].

Five of the studies that dealt with access for industry researchers were based on company surveys, using questionnaires, interviews or a combination of both. Three studies surveyed British companies, one in Denmark and one in Japan.

The UK's Publishing Research Consortium (alone in 2009 and in collaboration with UK's Joint Information Systems Committee (JISC) and Research Information Network (RIN) in 2011) commissioned two of the British studies. They were broad in nature aiming to compare access to research in four distinct groups: universities and colleges, medical schools and health providers, industry and commerce, and research institutes. However, both reports gave special attention to companies, especially SMEs. In the first study [5] data was collected from

several sources, e.g. lists of industrial/trade magazine subscribers as well as those who purchased papers through pay-per-view schemes. The respondents (1130 in total) represented different groups including 186 SMEs and 111 large companies. While 71% of SMEs (of those who need access) claimed to have easy access to the research literature, the number was much higher for large companies (86%) and universities (94%). They were also more likely (55%) to report facing access difficulties than large companies (34%) or universities (24%). They ranked open access as their third most common way to access literature, following personal and corporate subscriptions. Open access was ranked similarly by large companies, except that corporate subscriptions where more common than personal ones (unlike SMEs). Visits to local public library were the least common means of access, however 38% of SMEs and 35% of large companies reported using them.

The second study in the UK also confirmed the severity of the problem for SMEs, based on 2,645 survey responses from different types of institutions [42]. While no difference was found between the importance of research papers to SMEs compared to large companies, a larger proportion of the latter (78%) claimed to have easy access to the literature they need (compared with 69% of SMEs). Compared with only 44% of researchers in universities and colleges, 85% of industry and commerce researchers reported a recent access problem "that was not eventually resolved".

The third study in Britain was commissioned by JISC and draws on extensive interviews with representatives of 44 UK businesses, including 9 detailed case studies [80]. Authors admitted that these types of benefits are difficult to identify. This is partly because they couldn't find

cases of systematic open access usage by company researchers, i.e. open access was encountered "accidently" in the researchers' attempts to access literature. The study offered a useful typology of the different kinds of companies in relation to open access usage. It maintained that "research-oriented SMEs" would probably benefit from increased open access than other companies. However, it appears that for companies in general, the mere prevalence of open access literature is not enough to make use of them. Many lack "the skills and knowledge to develop and employ coping strategies when encountering paywalls". Also, some study participants claimed that any benefits for open access to industry are limited by the general irrelevance of academic research to industrial needs. For others, the benefits of open access were not in it being a knowledge transfer mechanism, but in being an efficient way to scan the large amounts of literature in order to identify potential collaborators from academia. This idea that companies use academic literature for purposes other than product development has been also discussed by Martin and Tang. They argued that humanities and social science research plays a role in providing companies with social knowledge necessary to solve "non-technical challenges that involve social choices" [24].

The Danish study also focused on SMEs, albeit on a smaller scale [43]. In addition to the online survey (which had 98 valid response), 23 interviews were conducted in attempt to provide a deeper analysis. Over half of the respondents claimed to experience difficulties accessing research papers. However, the importance they ascribe to such access is relative. Only 48% rated research papers as essential for their business, although the percentage was higher (64%) when only respondents in research roles are taken into account. Open access

journals and repositories as a means of access where rated above inter-library loans, local public libraries and pay-per-view, but still below in-house and personal subscriptions.

In Japan, two surveys conducted in 2008 and (a follow up) in 2011 showed an increase of the percentage of companies that could not introduce electronic journals citing budgetary constraints [81]. This might have been only because of rising journal prices. It is also possible that it resulted from the massive budgetary strains endured by Japanese companies after the Lehman Shock. It appears though that the problem might actually be more severe for smaller companies given that the number of those subscribing to 100 or more titles has increased to 16% from only 7% in 2008.

Pay-per-view access has always been a controversial issue in its own right, especially where individual paper prices are (as perceived by some) unjustifiably expensive or because of the uncertainty involved in judging about the paper's relevance just by reading the abstract. No wonder that 55% of customers of the article rental service *DeepDyve* come from organizations with less than 100 employees [82]. However, it appears that for companies the pay-per-view problem is two-fold, given that settling payments using credit cards is not their standard way of doing business. Of those who use pay-per-view in the Japanese survey, 30% faced difficulties related to making payments. This was also the second most common problem for industry and commerce researchers in the second British study. In this regard, it is important to note that pay-per-view was only used by companies as a mechanism to cope for lack of access, not as a standard way to access research [42].

One study sought to establish new indicators for the impact of open access outside academic circles. Bryan and Ozcan [83] tried to see if being open access enhanced an article's chance to be cited in patents. They matched close to three million patent applications with a set 132,872 papers published between 2005 and 2012 by 43 prominent biomedical journals. About half of the articles (54%) were freely available online (either through PubMed Central or publisher websites). Statistical analysis showed that open access articles received 28-59% more patent citations, although the authors warned that more investigation is needed to establish causality.

3.2.3 Policymakers

Evidence-based policymaking has been growing both as a field of research and in practice. Presumably, policy research units (usually operating on scarce funding) and individual policymakers can make good use of access to the widest range of academic literature relevant to their work. In fact, it is surprising that the earliest two studies in the set collected for this project sought to investigate the promise of open access in supporting evidence-based policymaking. Both studies were conducted by Willinsky in 2003 and 2004 with participation from Canadian policymakers. In the first one [84], 29 interviews were conducted with Canadian policymakers to better understand how they interact with research and what impact online (and open) access can have of this interaction. The study also offered a detailed discussion of the relevance of social science research to policymakers and the intricacies of the so-called evidence-based policy. Nwagwu and Iheanetu contributed a similar discussion in the context of their much broader survey of 121 Nigerian policymakers [85]. According to

their analysis, the availability of journals was not a good predictor of their usage. On the other hand, journal usage was significantly correlated to policymakers' educational qualification. In their conclusion, they emphasized the important role "meta sources" (e.g. bibliographic indexes) can play in enhancing policymakers' usage of different information sources. In this regard, and based on findings from his first study, Willinsky had developed a "Research Support Tool", which was meant to act as a companion to policymakers while they read policy-relevant research papers online. He tested the tool with 13 policymakers (also from Canada) and reported their views in his second study [86].

Later, a more extensive study was commissioned by JISC to investigate the possible benefits of open access for the public sector in the UK [87]. It attempted to estimate the direct and indirect benefits from open access to the public sector. Direct benefits where estimated in terms of cost savings. Based on survey data and other sources, they estimated that the existence of open access (both through gold and green routes) saves the UK public sector about 17% of what it annually pays to access its needed literature. One tenth of the saved amount is attributed to the more efficient use of time. The study concluded with a discussion of the indirect benefits of open access to the public sector (e.g. more informed decision making) recognizing the difficulties in quantifying such benefits.

In general, any positive impact for open access on the policymaking process should only be understood in the context of other limitations to evidence-based policymaking. For example, it is argued that decision makers in policy circles place might value information from someone they trust, than from a (generally) more credible source like scholarly journals [88].

3.2.4 Non-Profit Sector

The first of two studies to assess the needs and attitudes of NPOs regarding open access to research was a report commissioned by UK's JISC [89]. The study commenced with some "scoping interviews", then proceeded with a general survey and a detailed account of 10 case studies. The survey included 101 British charities with different sizes (with income ranging between less than £10K to over £10m). The majority of NPOs in the study (78%) reported that government-produced reports are the source of information they use most often, compared with 14% who claimed to use scholarly research most frequently. Still, 73% reported using journal articles and 54% used conference proceedings. While the response rate was small compared to the target sample of the survey (1,983 NPOs), let alone the total number of NPOs (132,962), the results align with findings from the interviews and case studies.

The majority of respondents (80%) cited the high cost as the major barrier to access research and 95% claimed that having research available for free online is the best enabling mechanism to increase their research use. It is also true that some reported asking student volunteers to provide them with papers using their university subscriptions. However, even in the case of larger NPOs (whose in-house libraries can afford subscriptions), the issue is further complicated by the fact that journals and academic databases are structured according to discipline. That might decrease the value of their subscriptions. For example an NPO serving people with a particular disability, "might be interested in research across a very wide range of disciplines, from health to sociology to engineering" [89].

In the second study [90], Moorhead and colleagues conducted an experiment with 92 researchers from a diverse set of nonprofits in the US. Two thirds (67%) of those given complete access (to the Stanford University Library collection of over 9000 health journals) have used it to view the full text of at least one article. On average, they viewed 2.2 articles per week and 3.7 abstracts for each viewed article.

Another related type of organizations is also important to consider here. In fulfilling their role as intermediaries, NPOs unions (e.g. National Council of Voluntary Organisations in the UK) or more sector-specific associations (e.g. Association of Medical Research Charities), also need access to a wide range of scholarly research. Their research digests and bulletins can significantly support their member NPOs by giving them "intellectual access" to the otherwise complex jargon of academic research [89].

3.2.5 Practitioners

The use of research by a few groups of practitioners has been studied specifically in the context of open access. Among different groups of practitioners, those in medical professions are thought to benefit most from open access to research publications. This is partly because their profession is perhaps more "science-intensive" than other professions and partly because biomedical disciplines have some of the most expensive journals among all. Medical practitioners are usually a standard example of open access beneficiaries outside academia (see Chapter 2), given the nature of their profession, where immediate access to information is suggested to have the ability to save lives. It is also true, however, that the information

needs of medical practitioners do not always align with what health researchers read [91] nor what they produce [92]. This is further complicated by the fact that in some cases (even when access is generally good) practitioners might not be aware of the opportunities offered to them [93]. The inability to locate information also poses the threat of assuming that it does not exist at all [91].

Although recently published [94], Spedding's review of evidence on whether "open access publishing facilitate the translation of research into health policy and practice" is based on publications listed in Bailey's 2010 bibliography [67] of literature on Open Access. This suggests that it might have missed important research on the topic published after 2010. Of the ten studies collected for this project (about the impact of open access on medical practice), only three were published before 2010. Nonetheless, the review offers very interesting insights into the issue. This includes suggesting that the reason why benefits to end users (e.g. practitioners) have been left out of the debate on open access is the fact that the open access movement itself has risen out of frustration with expensive journal prices. This consequently shifted everyone's focus to the issue of business models [94], which (among concerns about vanity publishing) has furthered the debate on the quality of open access publications.

Information-seeking behavior of medical practitioners has been a topic of interest for many researchers [91]. Of relevance here are those relevant to access to primary research literature. Based on a survey of 90 health personnel in the US, primary research was among their least common sources of information (used by 32% of respondents). They were motivated by the need to help a particular patient (35%) or a general interest in increasing their knowledge

(31%) [95]. In Ireland, a survey of health and social care professionals (HSCPs) showed a general lack of awareness about open access, which was mainly attributed to lack of time and institutional incentive to do research altogether. However, about 80% of them showed interest in doing research if given enough time and motivation [96]. The study also reported that over half of HSCPs who used documents from the Irish health repository (Lenus) have done so in practice for patient care, which supports the idea that open access to research can still have a great impact for those who are not generally active researchers. The problem seems to have existed for a long time because, as far back as 2005, 33% of family doctors surveyed by Andrews and colleagues claimed cost to be a barrier for the accessing literature they find relevant [97].

Being a well-defined group of people coupled with the relatively easy ways to reach them, health workers were the target of two experiments to study their relationship with open access. In the first experiment [92], Hardisty and Haaga showed that practitioners given a fulltext article were twice more likely to read it (and to some extent apply knowledge from it to a hypothetical patient) than those who were given only the citation. However, they also concluded that open access is a necessary (though not sufficient) reason for the diffusion of knowledge to the community of practitioners. This is a further proof of the importance of open access in the "chain of communication in health science" from researchers to health workers, especially in developing countries [98]. The other experiment [90] involved 336 practicing physicians from the United States (including a control group). The main conclusion was that practitioners consult the full text of the articles when given the chance. The experiment also suggests that OA-embargoes hinder the diffusion of research to practice

given that about half (49.9%) of the articles consulted by physicians were published within the previous year. The researchers conducted interviews for a subset of 38 physicians to gain more insight into the context in which these articles where used. The report for these interviews was published separately [99]. Regarding the importance of immediate access, Maggio and colleagues (who analyzed web log data of over 5000 health personnel working in the Stanford University Hospitals), that 20% of the research papers consulted by practitioners in 2011 where themselves published in 2011 [100].

Immediacy of access might also prove useful for another group of professionals. That is journalists. Although no formal study investigated this issue, preliminary analysis (reported recently in a blog post [101]) provides the interesting result that 60% of 11,523 research papers cited in news stories about cancer were behind a paywall, while 50% of the news stories where published within 2 weeks of publishing the research paper (25% within one day only).

Lawyers and Judges are another group that can make good use of access research (i.e. legal scholarship). Two studies tried to outline the efforts in this direction and provided and framework to understand the role legal scholarship can play in legal practice. Danner maintained that legal scholarship is more important for practitioners in civil law countries [102]. However, he cited another study that document cases where the Supreme Court in the United States (a common law country) resorted to citing academic literature in some difficult cases [103]. In the second study [104], Scherlen and Robinson provided a very sophisticated theoretical framework to make the case for open access based on principles of social justice

theory. In general, open access to legal research is still limited compared to other fields, despite the existence of initiatives like "The Durham Statement on Open Access to Legal Scholarship" [105] or the "Science Commons Open Access law Project" [106].

3.2.6 Patient Groups

Patients and patient groups have been frequently mentioned as an example of those who have a "right" to access research. In some countries, patient groups are very active organizations. Not only do they conduct advocacy and awareness-raising activities, but also fund and sometimes run research projects. One prominent example in this regard is PXE International, which was co-founded by a parents of children with a rare genetic condition called *pseudoxanthoma elasticum* (PXE). Sharon Terry, the nonprofit's founding executive director has explained [107] how she had to resort to illegitimate means to access enough literature to understand her children's condition, and how the existence of PubMed made it much easier later. Other than anecdotal evidence (like in Terry's case), there has been no formal investigation of the role open access can play is supporting the efforts of patient groups. A good starting point is papers like this one [108], signaling the different activities these organizations pursue and the extent to which they are supportive of open access research. Some even actively advocate open access. Close to one third of the 104 members of the Alliance of Taxpayer Access [109] are patient groups.

3.2.7 Independent Researchers

Independent (unaffiliated) researchers perhaps need open access to research the most. This is not because of their intensive research activities (unlike commercial researchers, who are probably the top beneficiaries of open access outside academia) but because they lack the means to get access to research otherwise [110]. It is also expected that in some fields, which have been historically pioneered by amateur researchers (e.g. anthropology, paleontology, etc.), the massive increase of freely available museum and archive collections on the internet can give momentum to for the return of the "armchair researcher" [111]. That is a researcher (like those prevalent in the 18th and 19th century), who without any professional credentials can conduct research based on their extensive library collections. Independent researchers also include those whose interest (or hobby) outside employment leads their research efforts [42].

In general, there are some ways independent researchers can access the research they need. For example, in 2010 16% of the requests handled by the document supply service of the British Library came from individuals (compared with 55% from academic sector, 17% from professional sources and 11% from businesses) [42]. These "hobbyists" can also visit public libraries or subscribe to an academic library as an external user. However, such schemes tend to be costly or at best impractical. They might need to travel for a long distance to reach a library with the needed resources, which can be challenging, especially for retired researchers. Also, these schemes do not always allow access to all resource because of library license restrictions [110].

3.3 Conclusion

This main purpose of this chapter was to explore the available evidence of societal benefits for open access research. It started with explaining the landscape of research about open access after which original research studies about societal impact were singled out. Despite being small in number, those studies revealed a variation of non-academic groups that can benefit from open access. The groups referred to were as follows. The total number below is 53 (not 52) because one study addressed two groups of open access users.

- society in general (13 studies)
- industry researchers (10 studies)
- medical practitioners (10 studies)
- patients (5 studies)
- policymakers (5 studies)
- unaffiliated researchers (3 studies)
- interested individuals (3 studies)
- nonprofit organizations (2 studies)
- law practitioners (2 studies)

These groups, in addition to others from the findings of Chapter 2, will act as the basis for the typology introduced in the next chapter.

4 Typology of Non-Academic Beneficiaries

This chapter presents a typology of the where access to research papers is needed along with a detailed description and justification for the typology. The proposed scheme is anticipated to provide guidance for future research on the issue.

- claims about open access benefits (Chapter 2)
- literature about the societal impact of open access (Chapter 3)
- review of literature from some disciplines that investigate various science/society interfaces (e.g. public understanding of science)

The main assumption of the typology is that usage of research papers outside of credentialed research establishments takes place in three different contexts. First, as is the case in academic research, non-academic researchers can require reading journal papers as a typical step in their research activities. Second, several professionals from time to time need to consult recent research papers, not to build on them with further research, but mainly to update their information and enhance their practice. Third, some individual citizens might have personal needs to satisfy through reading research papers. In all three cases, it might be the case that secondary sources (publications summarizing or synthesizing research papers) are produced in the process. These secondary sources, in turn, can extend this societal impact even further. The typology is explained by a diagram in Figure 4-1. In what follows, justification for each of the diagram components will be presented along with the argument

for how this typology can act as framework to guide future research into the issue of open access societal impact.

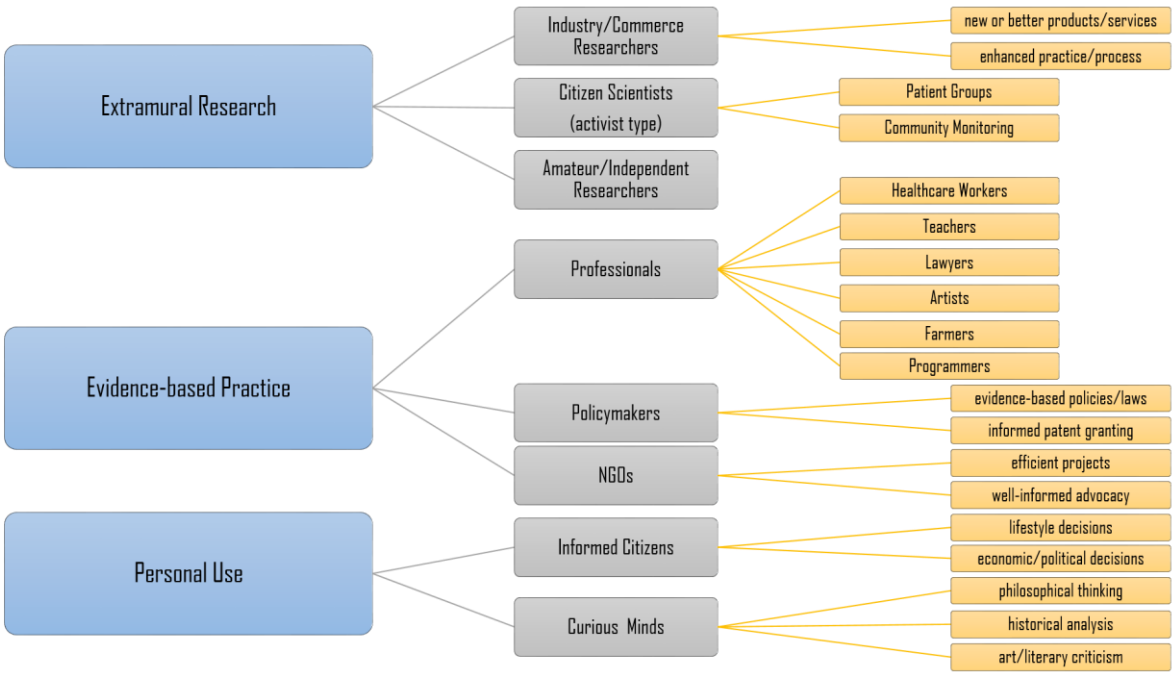


Figure 4-1: Typology of non-academic open access beneficiaries

4.1 Extramural Research

The term "extramural knowledge production" was explained by [112] as an umbrella term that involves all knowledge-producing activities that are conducted outside the university. Example for these are commercial research, citizen science, indigenous knowledge, amateur science, etc. Lave argued that despite the striking similarities among different forms of extramural research, they are rarely discussed together in the same context. This is also in part because researchers who study them come from different disciplines (e.g. anthropology, history of science, STS, innovation studies, etc.). This study assumes that, although with varying levels, researchers from different traditions of extramural research demonstrate a

common need to access scholarly resources that are available to their counterparts in universities. Three groups (of those mentioned earlier) stand out as examples of extramural research: industry researchers, patient groups and independent researchers.

In the diagram, industry researchers are displayed as a separate group of beneficiaries, albeit with the assumption that the category also includes non-manufacturing companies (e.g. financial sector companies). Patient groups are included under the larger group of citizen scientists. There has been many attempts to understand what citizen science is [113]. Many researchers in that field have also sought to classify the different activities that count as citizen science. They constructed their typologies based on the degree of participation of citizens, project goals or how projects start [114]. However, there is at least some consensus that citizen science projects are either driven by professional researchers (where citizens volunteer their labor to collect or process data) or designed and conducted by laypeople themselves (sometimes in cooperation with scientists). In the first case (volunteer type of citizen science) participation of lay people would normally be limited to the tasks assigned to them by the lead researchers. In the second case (activist type), lay people take the lead in designing and running the research process. One example for this second type is patient groups that conduct (not only fund) research projects with the participation of patients or their families. Another example is community monitoring activities that are geared towards environmental science research. This typology assumes that activist-type citizen science is one of the activities to which open access research is important. The third group of beneficiaries under this category are unaffiliated researchers. As mentioned above, both the phenomena of citizen science and unaffiliated researchers are much less understood than

industry research. Authors of the growing literature on both topics will hopefully extend their investigation to include the impact open access can have on these groups.

4.2 Evidence-based Practice

Despite the ongoing debate on whether open access articles get more citations than their pay-walled counterparts, we know that they at least get more downloads [115]. That is to say, some people, who might not necessarily be authors or researchers themselves, find open access articles to be good sources of information. Zuccala [29] suggests that the literature on information-seeking behavior is pivotal in understanding the impact of access to research outside academics contexts. Building on this argument, evidence for the impact of open access research can be sought by extending information-seeking behavior studies in different occupations. The typology presented here distinguishes three types of practitioners. The first group is people in different professions who require access to cutting edge information in their respective fields. These might include doctors, nurses, lawyers, engineers, farmers, artists, programmers or others. Research on evidence-based practice is an old tradition with its own dedicated books and journals. What would be important here is to study the extent to which access to primary research is essential for these professionals and devise new measures to measure it.

The second type of practitioners are policymakers (including both bureaucrats and decision makers). Towards the end of the 20th century, there was increasing emphasis on “evidence-based policymaking”. Examples of such trend can be seen in two different developments. On

the one hand, there was the establishment/expansion of many government units with mandates to collect information or analyze policy impacts. On the other hand, many universities and research organizations have initiated programs that try to support policymakers through activities like training bureaucrats or issuing policy briefs. A good example to such programs was the "consensus statements" issued by expert panels of the US National Institutes of Health to synthesize all available knowledge about specific topics in medicine. Research on evidence-based policymaking investigates the role of a wide array of evidence "types" (e.g. census data, polls, think tank reports, pilot projects, etc.) in the policymaking process. As shown earlier, a number of researchers have focused on the extent to which reading journal articles satisfies the information needs of policymakers. Nonetheless, it is important to warn that this argument assumes a direct link between acquiring policy relevant information and utilizing/applying such information to design or evaluate policies. This in itself is a controversial notion in the literature on evidence-based policymaking. While the rational choice theory requires that decision-makers need information about all available policy alternative before choosing a specific course of action, evidence have been building up to show that this is not the full picture. It is argued that bureaucratic (formal) structure and (informal) culture both play a key role in determining how (or whether at all) such information make their way to decision-making levels in government [88]. This issue must be closely considered during any effort to "measure" the impact of open access to research on policymaking circles. It is important not measure this impact, but to study the different roles research evidence plays in the policymaking environment.

Those working for the nonprofit sector are third group of practitioners relevant to this typology. Access to scholarly research can be of value for a non-profit organizations for two main reasons. One reason is the need for sound information to be used in the NPO's advocacy activities, that being anything from lobbying in parliaments to awareness campaigns on the street. Reliable information from the results of peer-reviewed research can act as a strong evidence-base to structure claims and arguments, especially when justifying the request/use of funds. This can be relevant to organizations like environmental advocacy groups, for example. The other reason for the importance of research results is their usage in designing and implementing projects and initiatives. An NPO that is well-informed about all studies and reports in its field of work is in better position to design relevant projects and to successfully implement them. An NPO in the field of social work might, for example, use the results of research on homelessness to establish facilities or services that are relevant to the needs of homeless people. The body of research on the use of evidence by nonprofits is not as large as the previous two groups of practitioners. More surveys and experiments can prove useful in this regard to measure the impact of open access.

In all cases, any discussion of evidence-based practice should take into account factors other than the availability of the required information. For example, the ability to put problems faced in practice in terms of questions to research must be taken into consideration. The same applies to the ability to judge on the quality of the consulted research papers and to draw practical implications from them [93].

4.3 Personal Development

Common sense would suggest that laypeople would neither have the interest, nor the skills to consult scholarly research papers to fulfill their information needs. The implication for this is that the impact of open access to research in this case is either nonexistent or minimal enough to ignore. The *D.C. Principles for Free Access to Science* (a document well-known for its advocacy against open access) questioned why "the NIH thinks technical medical information should be offered up to the public in so cavalier a fashion ... It is debatable whether members of the general public can actually benefit from reading the original research literature, as its arcane and specialized reporting is intended primarily for other researchers". This position can be contested based on studies from the field of Public Understanding of Science (PUS), where a general increase in lay people's comprehension and engagement with scientific issues has been documented over the last few decades [116]. Such increase has been attributed to a variety of factors, not the least being the global increase in the amounts of people receiving higher education. While PUS statistics are mainly concerned with natural sciences, it is not difficult to assume that similar effects are present with regards to arts, humanities as well as social sciences (AHSS). It is also important not to forget that researchers from many AHSS disciplines have historically been able to engage more lay people with their work through writing books, which has not been the case with their natural science counterparts. Zuccala and colleagues [117] have shown that some history monographs that are widely cited in academic journals still appeal to many lay readers, as measured by the reviews they give on the platform Goodreads. On the other hand, the

unprecedented growth in the availability of MOOCs and other Open Educational Resources (OER) can definitely help in preparing interested lay people to handle research papers whenever the need arises to consult them.

Personal use is probably the least understood category of usage of research papers. This study assumes that attempting to classify the different types of this usage will provide much-needed guidance to future research efforts. Based on previous studies of the impact of open access on laypeople's lives, two distinct types of usage can be distinguished. First, there is the activity of consulting research papers to inform one's behavior either in private (lifestyle) matters (e.g. parenting, nutrition, coping with illness) or in public matters (e.g. following economics or public policy research) in order to actively fulfill one's role as an informed citizen. The overall sentiment of laypeople who took part in Zuccala's focus groups was that that they "tend to be positive about the fact that free scientific/scholarly information has potential to empower them, especially in terms of personal choices and personal decision-making" [18].Second, there is reading research with the aim of satisfying one's curiosity (e.g. history, philosophy, art, religion etc.). The impact of this second type on society is two-fold. Directly, it can provide personal enrichment and help people fulfill higher purposes in life. It also has an indirect effect on the development of different cultural industries (e.g. galleries, tourism, books, etc.) and other the flourishing of other cultural activities.

4.4 Considerations in Designing the Typology

In essence, these three categories (extramural research, evidence-based practice and personal use) all assume that research papers are "sought" by members of society who need them. This study argues that an open access mode of scholarly publishing would make the task of consulting research easier for everyone (not just within the academic community). Indeed, some have argued that unlocking the body of scholarly literature is in itself a virtue, regardless of what utility it holds for society. Such position also assumes that new "uses" of research would arise if we just focus on making it available. However, this study chose to emphasize already existing needs and lay out the framework to better understand them.

It is true that for many large corporations, some think tanks, NPOs and even individual practitioners and lay persons the current subscription prices are affordable at least for the journals they need. This, however, does not eliminate the need for open access for the obvious reason that the system in this case will be biased towards those who are wealthy enough to afford the rising prices. There is another reason too. That is, as we move to the Mode 2 of knowledge production [118], much of the research becomes trans-disciplinary and it will no longer be possible for a research organization to stay bound to a specific set of journals to cater for their needs.

It is also important to note that overlap between the three categories of the proposed typology is inevitable. For example, personal use might in itself be a form of practice, in the sense that individuals read the research to apply it in their own lives (e.g. parents reading research on

child psychology). However, it was better to separate both categories owing to the obvious difference in the metrics we need to evaluate them. In other words, surveying doctors to know the number of cases they needed to consult research papers to treat, would be a lot different from measuring the impact of philosophical research on creating a rational public. The other reason for separating the two is the more direct impact professional practice has on the economy, which might be a variable of interest to many.

Another area of overlap also exists between extramural research and practice. An example for this would be an R&D department in a company that directly applies part of the research its consults (e.g. research on managing R&D departments) and builds on the other part with more research (i.e. research on the products it works to develop). In such case, measurements of the research production of a firm (e.g. number or articles/patent applications) will not be an accurate reflection of its open access research dependence. Similarly, a consulting firm might not necessarily produce any written documents for a given project, although its main aim is to do research. Practice and personal use also have their area of overlap, for example, in the grey zone between a patient's own search for information and his/her patient group advocacy efforts.

"Open access" to research for the purposes described in this paper does not always mean free/gratis access. It is true that in some contexts (perhaps for large corporations or science-intensive SMEs) mining of a large number of research papers might be necessary. In this case, to achieve full utility, those papers need to be published under licenses that would allow such activities. Another context that dictates expanded reuse rights is the case of practitioners who

might need to extract and remix information from different sources to develop practice guidelines (clinical protocols is one example). A permission to print (and maybe distribute on a small scale) the research paper might also be necessary in the case of citizen science projects.

4.5 Conclusion

The typology presented in this chapter categorizes non-academic users of academic research according to whether they use it for more research (Section 4.1) or for application. Applying research from journal articles can either be in the context of their professional (Section 4.2) or personal (Section 4.3) lives. Each of the examples of users under each category can be the topic of research in the future. Some of these groups have been repeatedly studied (e.g. industry researchers) and some of them not studied at all (e.g. unaffiliated researchers). The end of this chapter marks the fulfillment of one of the two main purposes of this dissertation, which is identifying both the current and the potential beneficiaries of open access outside academia by presenting the typology above (i.e. it answers Research Question 1). The next chapter will try to fulfill the other purpose by suggestion new ways to study the impact of open access on two groups of users in particular.

5 Enhancing the Evidence Base

The previous three chapters represented the endeavor to collect and synthesize evidence on the issue of open access societal impact, as well as to develop a framework to use for future research on the issue. What this chapter tries to achieve is to report on two attempts to create evidence on how open access can impact two of the groups mentioned in the typology, namely industry researchers and unaffiliated researchers. Both groups (as part of the Extramural Research category in the proposed typology), are important to understand because they are by definition the top beneficiaries of access to scholarly journals. As is evident from the findings of Chapter 3, several studies have looked into usage of academic research among industry researchers and the possible implications open access can have in this regard. This is unlike the case with independent researchers where only anecdotal evidence is available. In the first section of this chapter, a new method of investigating the impact of open access on industry researchers is presented, along with the results of using it to analyze a dataset of 1104 pharmaceutical company patents from the United States. The pharmaceutical industry was chosen as an example of a "knowledge-intensive" industry that ascribes a lot of importance to relations with universities and usage of scholarly journals than other industries, according to the study by Cohen and colleagues [44]. The second section of this chapter sheds light on the issue of unaffiliated researchers and how open access research can affect their publication activity (measured by publications indexed in Scopus). They were chosen mainly because of they are the least group among those in the typology whose information-seeking behavior is understood. Preliminary analysis for their citation behavior is provided along with directions for future researchers on how to expand on it.

5.1 US Pharmaceutical Companies

Just like growth theories of the 1950s which utilized the long history of collecting and organizing macroeconomic data and indicators, the emergence of the so-called "knowledge economy" has called for new types of data and indicators that can make it easy to quantify and measure intangible resources like knowledge, ideas and technology [119]. At the crossroad between the field of Innovation Studies and Bibliometrics, "science linkage" emerged to fulfill this purpose. This research field sought to quantify and measure the impact of scientific research on innovation and (indirectly) on economic development. The main idea was using the rich patent data (mainly from the United States Patent and Trademark Office, USPTO) including the citations they have to journal papers to construct different indicators that would enhance our understanding of knowledge flow and spillovers. The seminal works of Kuznets [120], Griliches [121], Narin [122] and the more recent work by Jaffe and Trajtenberg [119] have all paved the way for the sophisticated scholarship we have today in this field. Following this tradition, citations in patents (as opposed to citations in journal articles written by industry researchers) were analyzed in this study in attempt to tackle the issue of open access from a science-linkage / knowledge transfer perspective.

Although there has been one attempt to use patent analysis for measuring the impact of open access research (details in Section 3.2.2) which apparently took place in parallel with the one presented here, this study takes a different approach from the one used by Bryan and Ozcan. Both studies assume that patent analysis would eliminate the bias of self-report generally experienced in surveys. Also, they both focused on biomedical research because of the strong

relation it has with the biotech industry in general and the pharmaceutical industry in particular. The main difference between the two studies is that this one, instead of investigating the issue from the side of journal article citations and impacts, focuses on the users of these papers (i.e. companies that own the patents).

Building on evidence about the negative impact of high journal prices on SMEs (also discussed in Section 3.2.2), the main idea of this study is attempting to understand how the company's size relates to its usage of open access research [123]. The main hypothesis is as follows:

Small companies use more open access research than large companies

In order to test this hypothesis, a dataset was constructed by matching citations in patents of US pharmaceutical companies to the sizes of those companies at the time of patent examination (usually assumed to be one year before the patent grant year). The assumption is that the company size is a measure for its ability to subscribe to journals. Hence, small companies are expected to cite more open access journals in the patents than large companies because of their inability to subscribe to all the journals they need.

5.1.1 Data

5.1.1.1 Sources

For the purpose of this study, a dataset was constructed by matching records from two different sources. The first source of data is a list of companies and their financial information obtained from Bureau van Dijk's Orbis database. Among over 140 million companies in the

database, the selection was restricted to US-based companies working in the pharmaceutical sector (NACE Primary Code 21). The initial set included 1109 companies. This number excludes 338 companies that were designated as "very large" by the Orbis database. They were excluded for practical reasons given that they own enormous amounts of patents (which would have significantly complicated the analysis process) while at the same time they were not as relevant to the purpose of the research as other categories of companies were. Orbis database contained financial information about companies in the decade between 2004 and 2015. Therefore, in preparation for the later step where regression analysis is conducted using matched records of patent and financial data, companies which own no US patents (over the same 10-year period) were excluded. At this point the dataset included 654 companies owning a total of 9221 patents (Table 5-1).

Table 5-1: Counts of companies, patents and citations at different stages of data refinement

Description	Company Count	Patent Count	Citation Count
All US Pharmaceutical Companies (excluding "very large")	1109	13534	---
…… with US-registered patents between 2006 & 2015	654	9221	---
…… with matching financial data (1 year before patent grant)	354	1104	39037
…… that cite academic journals	280	825	28694
…… that cite Open Access journals	193	483	23476

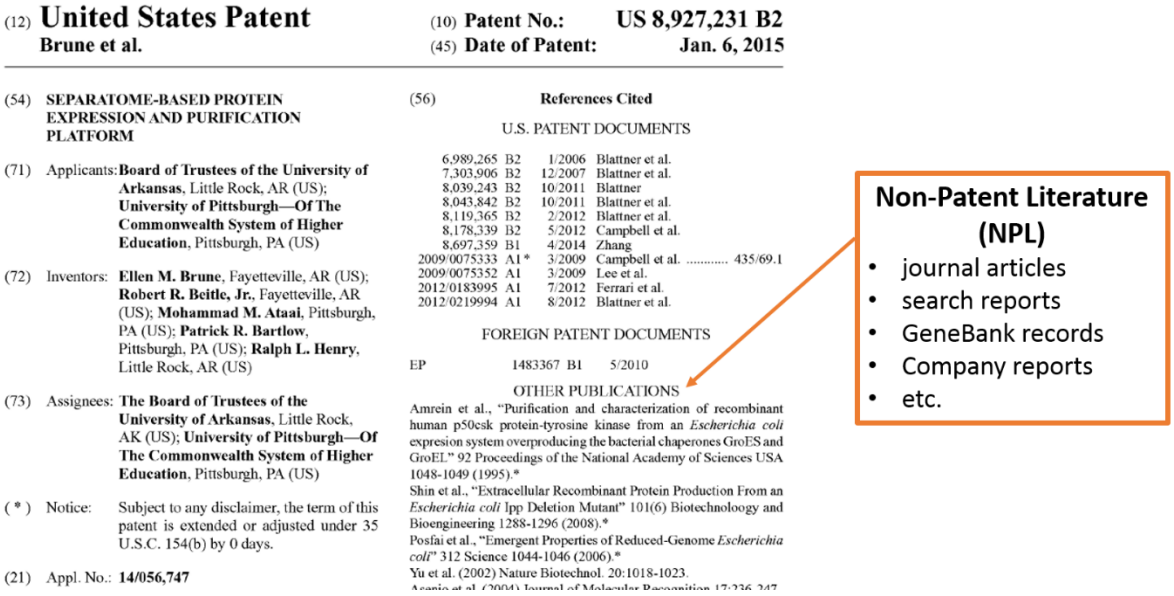
For each company, the data retrieved was as follows: company name, owner group (if any), number of employees (2005-2014), total number of owned patents in addition to the classification and the identifier for each one of these patents. Only granted patents were taken

into consideration, other types of patent documents (e.g. patent applications) were excluded. Also, only patent documents listed in Orbis database were included for each company. No independent check was made to verify whether the company has more patents. However, the USPTO database was used to confirm that patents mentioned in Orbis are actually owned by their respective companies.

Because of the restriction that patents in the dataset should be matched to company size information one year before patent grant date, patents owned by companies where no corresponding financial data was available had to be excluded. This turned out to be the majority of patents. Out of 9221 patents, only 1104 could be matched. Section 5.1.1.2 below discusses the issue of missing financial data in more detail. Also, in trying to remedy this problem, an attempt was made to interpolate some of the missing data. More details about the interpolation process can be found below in Section 5.1.1.3.

The second source of data was the journal citations in patents collected from the United States Patent & Trademark Office (USPTO) database. Patent identifiers from the Orbis data were used to collect more information about the patents. A custom-built Microsoft Access Form was used import data from USPTO website and to record the required data in a specially designed database. Figure 5-2 shows a screenshot of the form in operation. Data on patent grant year and owner company names were confirmed to match those extracted from Orbis. In addition, two other types of data were collected. These are co-owners of the patent and citations to non-patent literature (NPL).

Figure 5-1 shows the first page of a typical USPTO patent document. Its "References Cited" section includes citations to US patents, foreign patents and "other publications", which is usually referred to as non-patent literature. This NPL includes anything from citations to journals articles to letters of communication during the patent examination process. NPL citations (starting in year 2001) are marked to identify whether the citation was made by the applicant or by the examiner during the examination process [124]. Only applicant-provided citation were considered for the purpose of this study.



(12) **United States Patent**
Brune et al.

(10) **Patent No.: US 8,927,231 B2**
(45) **Date of Patent: Jan. 6, 2015**

(54) **SEPARATOME-BASED PROTEIN EXPRESSION AND PURIFICATION PLATFORM**

(71) Applicants: **Board of Trustees of the University of Arkansas**, Little Rock, AR (US); **University of Pittsburgh—Of The Commonwealth System of Higher Education**, Pittsburgh, PA (US)

(72) Inventors: **Ellen M. Brune**, Fayetteville, AR (US); **Robert R. Beitle, Jr.**, Fayetteville, AR (US); **Mohammad M. Ataai**, Pittsburgh, PA (US); **Patrick R. Bartlow**, Pittsburgh, PA (US); **Ralph L. Henry**, Little Rock, AR (US)

(73) Assignees: **The Board of Trustees of the University of Arkansas**, Little Rock, AK (US); **University of Pittsburgh—Of The Commonwealth System of Higher Education**, Pittsburgh, PA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **14/056,747**

(56) **References Cited**

U.S. PATENT DOCUMENTS

6,989,265	B2	1/2006	Blattner et al.	
7,303,906	B2	12/2007	Blattner et al.	
8,039,243	B2	10/2011	Blattner	
8,043,842	B2	10/2011	Blattner et al.	
8,119,365	B2	2/2012	Blattner et al.	
8,178,339	B2	5/2012	Campbell et al.	
8,697,359	B1	4/2014	Zhang	
2009/0075333	A1 *	3/2009	Campbell et al.	435/69.1
2009/0075352	A1	3/2009	Lee et al.	
2012/0183995	A1	7/2012	Ferrari et al.	
2012/0219994	A1	8/2012	Blattner et al.	

FOREIGN PATENT DOCUMENTS

EP	1483367	B1	5/2010

OTHER PUBLICATIONS

Amrein et al., "Purification and characterization of recombinant human p50csk protein-tyrosine kinase from an *Escherichia coli* expresion system overproducing the bacterial chaperones GroES and GroEL" 92 Proceedings of the National Academy of Sciences USA 1048-1049 (1995).*

Shin et al., "Extracellular Recombinant Protein Production From an *Escherichia coli* lpp Deletion Mutant" 101(6) Biotechnoloogy and Bioengineering 1288-1296 (2008).*

Posfai et al., "Emergent Properties of Reduced-Genome *Escherichia coli*" 312 Science 1044-1046 (2006).*

Yu et al. (2002) Nature Biotechnol. 20:1018-1023.

Non-Patent Literature (NPL)
- journal articles
- search reports
- GeneBank records
- Company reports
- etc.

Figure 5-1: Front page of a typical USPTO granted patent document, with the illustration of the location of journal article citations

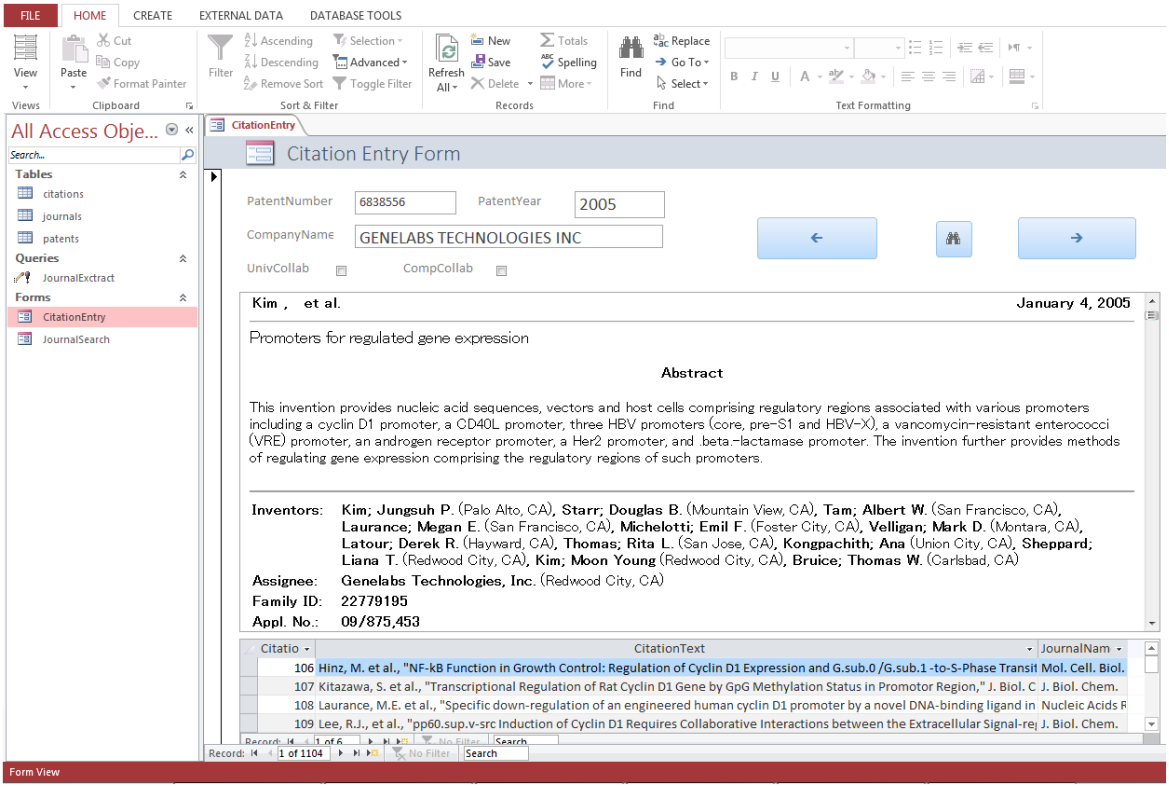


Figure 5-2: Screenshot of the Microsoft Access form designed to collect the citation data

After extracting NPL citations for all patents in the dataset (39,037 citations), keyword search was used to exclude the majority of citations to non-journal publications (e.g. letters, textbooks, reports). Those missed during this process were excluded during the following stage. At this stage, computer-assisted manual search was used to extract journal names from the remaining about 30,000 (mostly journal) citations. NPL citations are not structured and each citation is provided in free text format. The problem is further complicated by the fact that this free text does not follow a specific citation style, meaning that applicants had the freedom to describe the cited paper in any format they want. This is in addition to different ways to abbreviate journal names and non-standard uses of punctuation marks. All of this made it difficult to use citation parsing software (assuming the results will not be reliable and

will need further manual confirmation) and necessitated the tedious task of manual extraction of journal names [125].

Journal names could be identified within 28,694 citations to journal articles. Also, 179 patents were excluded from the dataset, where no citations to academic journals existed. In the remaining 825 patents that included citations to journals, citations where identified to be to open access journals through online investigation of the journal history (which in some cases included looking at historical versions on the journal website using the Internet Archive's Wayback Machine). Current open access journal databases (e.g. DOAJ) could not be used to determine if the journal is open access because what matters is whether the journal was open access at the time the article was published. A set of 79 citations (0.2% of the total) could not be identified. They were considered non-journal citations and excluded from further analysis.

Citations where the applicant specifically mentioned that they cite the paper abstract, or those where an abstract service is cited (e.g. Chemical Abstracts) were considered to be citations to open access papers, given that all online journals make their abstracts accessible at no cost. These accounted for the majority of open access citations (1330 citations, 61% of open access citations and 4.6% of total journal citations). Therefore, as a robustness check, regression analysis was repeated after excluding citations to abstracts altogether (see Section 5.1.3.2).

5.1.1.2 Patterns of missing data

As noted above, the majority of companies had gaps in their financial data (i.e. no reported company size for one or more years). It was necessary to understand the pattern (if any) of this missing data and attempt to understand why it was missing.

Table 5-2: Distribution of missing Employee Count Data in Orbis database. The maximum number of missing data points is 10 as the database covers 10 years (2005-2014)

Years without Data	Count of Companies
0	1
1	1
2	5
3	19
4	40
5	99
6	144
7	132
8	80
9	73
10	60
Total	654

Of a possible total of 6540 data points (654 companies with 10 data points each), only 2132 (33%) were available in Orbis. The distribution in Table 5-2 shows the large number of missing data points. Three quarters (74.7%) of the companies had 5 or more data points missing, which suggests that applying different time-series data interpolation methods (e.g. cubic spline interpolation) would be unlikely to give usable results.

Despite the fact that the majority of patents were excluded in matching step, the remaining dataset still contained over half of the companies (54%). However, checking the representativeness of the remaining companies could not be done at the company level.

Because it was panel data, each company had more than one size reported (for different years). Orbis provided categories for company size (which were used earlier to exclude "Very Large") companies. However, these could not be used to check sample representativeness because they are calculated using only the latest available financial data from each company (not always 2014). This does not reflect changes in company size over the included 10-year period. Therefore, in order to check the sample representativeness records for individual years were used as if they were separate records (for different companies). In Table 5-3 the percentage of records belonging to a particular size category are reported for all of the raw data from Orbis (left column), all data points that could be matched to patents (middle column) and all matched data including interpolated data (right column). The following section (Section 5.1.1.3) explains the interpolation process further.

Important to note here is the fact that the raw data presented in Table 5-3 is only 33% of what could have been available, had all companies submitted their data for all of the entire 10-year period. Therefore, any check of sample representativeness is not actually done against the whole population of US pharmaceutical companies, but only against the proportion of information available from the Orbis database.

To understand the reason behind this large number of missing data points, there was an attempt to contact Bureau van Dijk (the company that own Orbis) via email to inquire about their data collection method. It was sent on June 30, 2017, but there was no reply. Further analysis of the missing data was done to speculate what the reason might have been. Figure 5-3 below shows a growth in the percentage of available data over the whole 10-year

period. Most of the missing data is from earlier years. The reason for missing data might have been due to inefficient data collection mechanism that have been improved over time.

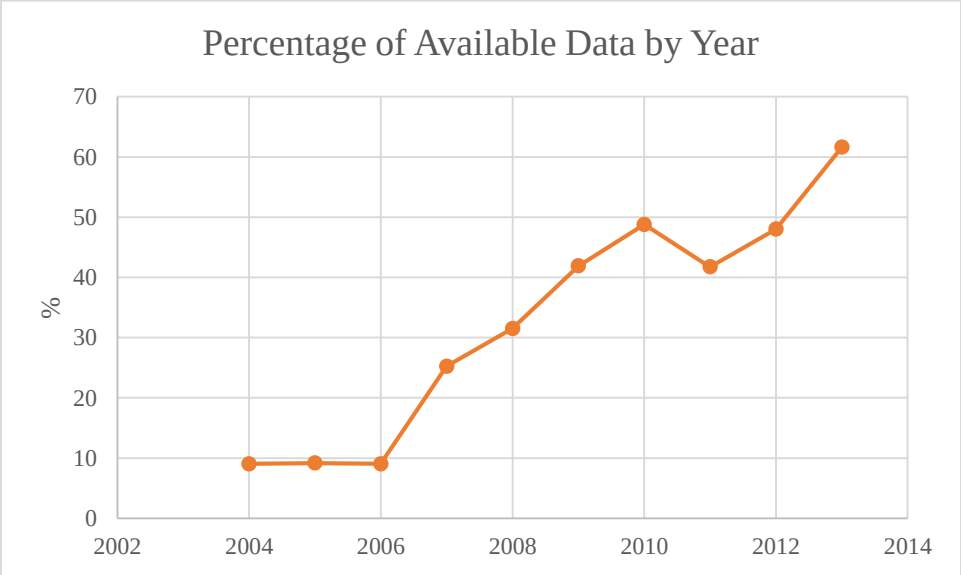


Figure 5-3: Decrease in missing data over time

5.1.1.3 Interpolation of missing data

An attempt to get around the problem of missing data was done through imputation. The idea was to construct a number of company size data points (as measured by number of employees) using already available data and then match them with patents owned by these companies (granted one year after the year for which the company size was constructed). Of the 654 companies in the data set, 212 had cases where the number of employees was missing for one year but available one year before and one year after. All of these cases existed only between the years 2008 and 2012, which supports the assumption made earlier about the reason behind missing large portions of data. For these cases, the arithmetic mean was calculated and taken to be the company size at the year in the middle. A total of 249 data

points were add (for some companies, there was more than one of these cases). Only 102 companies owned patents that could be matched to the constructed data points. Matching those resulted in 153 new records added to the original 1104 patent-company size pairs used in this study. In short, the imputation process resulted in expanding the original dataset by about 14%. Among these 153 patents, 93 had at least one citation to a journal article and were therefore added to the regression. Table 5-3 below shows how adding the interpolated data enhanced the sample's representativeness.

Table 5-3: Percentage distribution of company sizes in the raw data, before interpolation and after interpolation

Number of Employees	All Data from Orbis (n=2132)	Matched Data (n=825)	Matched Data including interpolation (n=918)
1 - 9	39.7%	16.6%	17.8%
10 - 49	27.3%	22.2%	23.0%
50 - 249	25.6%	43.3%	41.8%
250+	7.3%	17.8%	17.3%

5.1.2 Variables

The unit of analysis in this study is the individual patent. Each row includes information about the patent itself, citations from its NPL section and information about its owner company. Even if several patents were owned by the same company they still constituted distinct record because the financial data for each company change year by year (matching the patent grant years). Detailed description of the different variables used in the regression is given below.

5.1.2.1 Dependent variable

A count model was used for the regression (more details in Section 5.1.3). The dependent variable (*OAjournals*) is the count of citations to open access journals mentioned in the patent document.

Based on the process of investigating the status of all cited journals and their histories (which included in some cases consulting the Internet Archive for earlier versions of the journal website), citations where split into three categories:

- citations to journals that were established as open access or "born open" (141 citations, 6.5% of open access citations and 0.5% of total journal citations)
- citations made *before* their journal converted to open access (486 citations, 22.3% of open access citations and 1.7% of total journal citations)
- citations made *after* their journal converted to open access (220 citations, 10.1% of open access citations and 0.7% of total journal citations)

Citations from all of these three categories where considered open access given that the majority of journals that convert to open access make their archives openly available. Citations to abstracts were considered citations to open access papers in the one set of models and were dropped altogether in the another set.

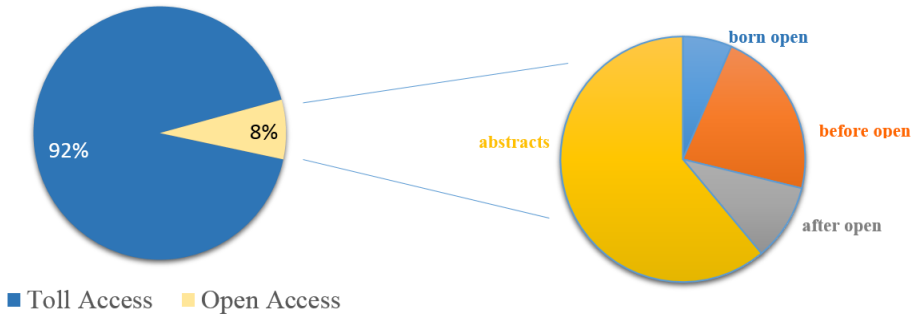


Figure 5-4: relative proportions of different types of citations that were considered open access: citations to abstracts (1330 or 61%), citations to born open access journals (or), citations to articles published before the journal converted to open access (486 or 22%) and citations after conversion (220 or 10%)

Before interpolation, there was a total of 2,177 open access citations making up about 8% of all journal citations (Figure 5-4). The ratio was not much different after adding open access citations from patents analyzed after interpolation (Table 5-4).

Table 5-4: Description of data added through interpolation

	Before Interpolation	After Interpolation
All Patents	1,104	1,257
All NPL Citations	39,037	45,042
Patents Citing Academic Journals	825	918
Citations to Academic Journals	28,694	33,216
Open Access Citations	2,177	2,532
Open Access Citations (excluding abstracts)	847	944

The distribution of citations both before and after interpolation is very skewed, with most patents not citing any open access journal. Figure 5-5 shows the distribution open access citations including those added after interpolation and Figure 5-6 shows it for all academic

journal citations, open access or not (also including citations added after interpolation). This skewness is part of the reason why a zero-inflated regression model was used.

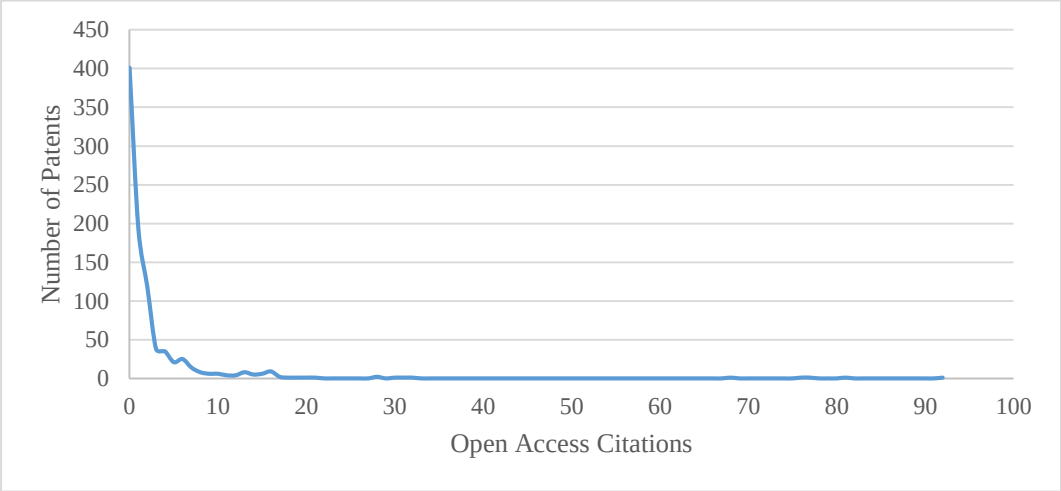


Figure 5-5: Distribution of open access citations over patents in the dataset (including interpolated data)

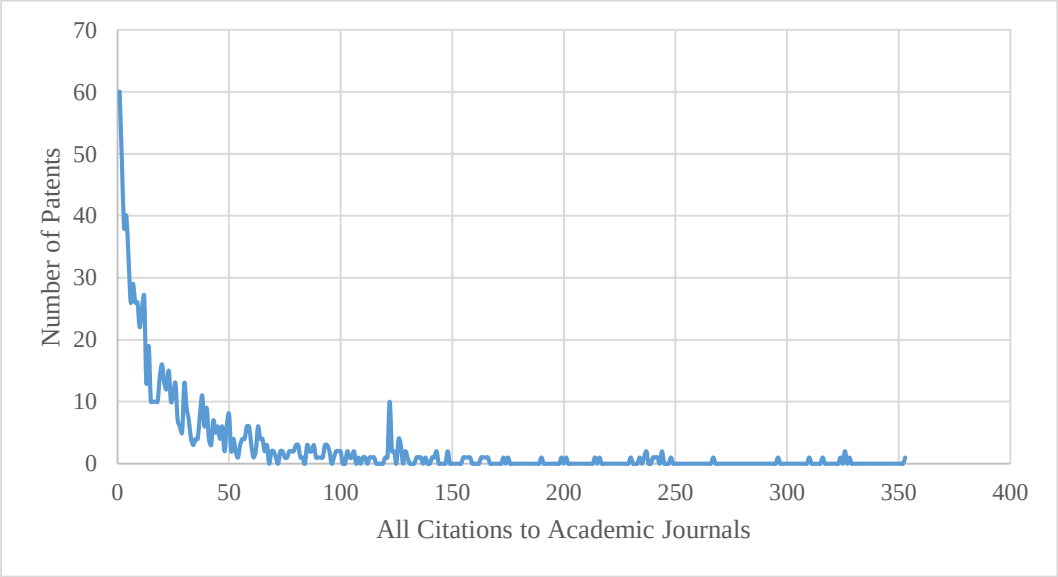


Figure 5-6: Distribution of all journal citations over patents in the dataset (including interpolated data)

5.1.2.2 Independent variable

The independent variable in this regression analysis is the company size as measured by its number of employees. The range of sizes included in the dataset covers all but two of the nine company size categories proposed by the Frascati Manual [126]. Only companies with 0 employees or those with more than 5000 are not represented in this dataset.

5.1.2.3 Control Variables

Another set of variables were assumed to influence which journals are cited in patents of pharmaceutical companies and consequently were included in the regression as control variables. First, the total number of cited academic journals had to be accounted for because some patents tend to cite academic journals less than others depending on the novelty of the technology as well as other factors. Second, the overall technological capability of the company (measured using its patent profile) is a good determinant of its research activity. Research active companies tend to subscribe to more journals than other companies, which means they would rely less on open access. Third, although all companies in the dataset are pharmaceuticals companies, not all patents are classified as such. Many pharmaceutical companies own patents that belong to other fields of research (e.g. polymers). In general, if the company allocates resources to journal subscriptions, the assumption is that it will give preference to biomedical journals. This means that the company researchers are more likely to seek open access journals when their invention is outside the main field of the company. Fourth, a given company might afford to subscribe to a limited number of journals. However, if the patent at hand was a result of research collaboration, it is possible that sharing journal

access privileges with collaborators would decrease the need to seek open access alternatives. This effect is accounted for by three different control variables, namely, collaboration with university, collaboration with another company and access to resources owned by a parent company. Table 5-5 below gives the operational definitions of each of these variables.

Table 5-5: Operational definitions of control variables

Control Variable	Corresponding Indicator
Total Cited Journals (ALLjournals)	total number of academic
Technological Capability (*TechCap*)	total number of patents (including patent applications) affiliated with the company divided by its number of employees
Patent Relevance (*RelatedClass*)	equals 1 if the patent is classified under A61, C07,C08,C12 (CPC System) and zero otherwise
University Collaborator (*UnivCollab*)	equals 1 if the company co-owns the patent with a university and zero otherwise
Company Collaborator (*CompCollab*)	equals 1 if the company co-owns the patent with another company and zero otherwise
Group Membership (*GroupMember)*	equals 1 if the company is owned by a parent company and zero otherwise

Finally, year dummies were essential to add given that the body of openly available literature has been growing ever since the Open Access movement emerged. This is evident from the increasing number of open access journals listed in the Directory of Open Access Journals (DOAJ) [127] as shown in Figure 5-7 below and the increasing popularity of open access policies as previously demonstrated in Figure 1-3 (see Section 1.4). The assumption is that the more open access articles become available, the more companies use of them.

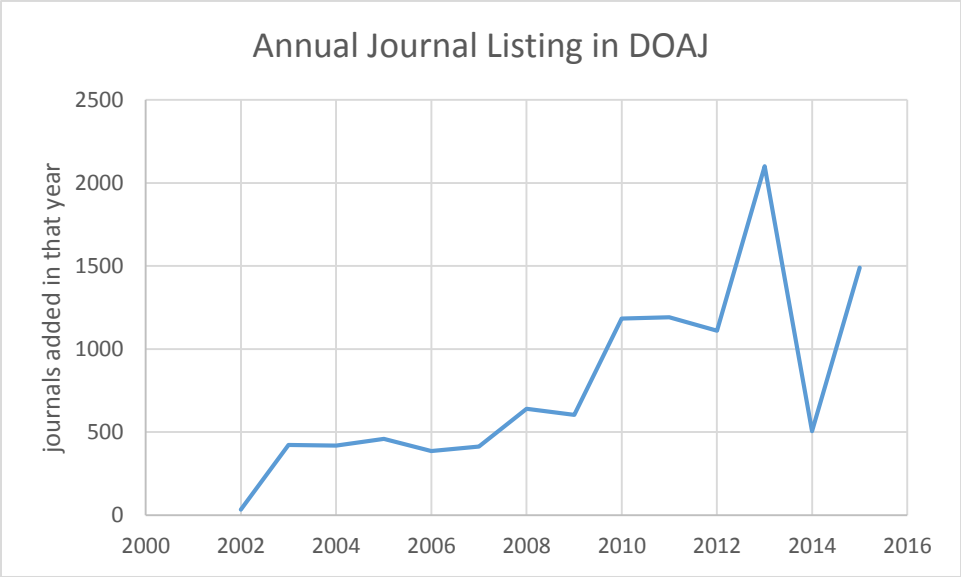


Figure 5-7: Annual journal listing in DOAJ

5.1.2.4 Correlation among Variables

The correlation matrix including all regression variables (except year dummies) is shown in Table 5-6. Despite being small in value, the correlation coefficient of *OAjournals* (the dependent variable) and *EmployeeCount* (the independent variable) has a negative sign, which agrees with the main hypothesis that the larger the company, the less open access it uses. The same applies to assumptions about technological capability, group membership and collaboration with university, which all had negative correlation coefficients as expected. A somewhat stronger correlation exists between the usage of open access and citing journal articles in general.

Table 5-6: Correlation matrix among all regression variables (excluding year dummies)

	OAjournals	ALLjournals	EmployeeCount	TechCap	Group Member	Related Class	UnivCollab	Comp Collab
OAjournals	1.0000							
ALLjournals	0.4486	1.0000						
EmployeeCount	-0.0505	-0.0006	1.0000					
TechCap	-0.0116	-0.0248	-0.0408	1.0000				
GroupMember	-0.0111	-0.0432	0.0189	-0.0112	1.0000			
RelatedClass	0.0181	0.1008	-0.0080	-0.1014	-0.0073	1.0000		
UnivCollab	-0.0202	0.1091	-0.0725	0.0039	0.0128	0.0611	1.0000	
CompCollab	0.0008	-0.0113	0.0939	-0.0104	0.0301	0.0034	-0.0496	1.0000

5.1.3 Analysis Results and Discussion

Owing to the large number of zero values in the dataset, it was decided to use a regression model that can accommodate this prevalence of zeroes. As mentioned above, the zero-inflated negative binomial was used. It was also a better fit to the skewness found in the data (Figure 5-5). Initially, a Negative Binomial model was used, but applying the Vuong test showed the zero-inflated model was more suitable. Earlier in the analysis a likelihood ratio test rejected the hypothesis that the over-dispersion coefficient (alpha) was equal to 0 and ruled out the option to use a Poisson model.

Two tests for heteroskedasticity were conducted. Result from the Breusch-Pagan test showed that the heteroskedasticity could be detected in the dataset, while White test failed to reject the hypothesis that the data is homoscedastic. Regression results reported below are based on robust standard errors.

Regression took place using four different specifications. In Model I-1 and Model I-2, the definition of the dependent variable (*OAjournals*) included counting citations to abstracts as citations to open access papers. Model I-1 was the baseline model (with control variables only) and Model I-2 included the independent variable (*EmployeeCount*) in addition to the control variables. The same was true for Model II-1 and Model II-2 except that abstracts where dropped from the dataset, just like all other citations to documents other than academic journals (e.g. patent search reports, magazines, company brochures, etc.). There was a slight decrease in the number of observations due to the existence of patents that only cited abstracts.

Analysis using these four specifications were all repeated but using the original data in addition to the interpolated data. Table 5-7 and Table 5-8 present the results from all regressions.

5.1.3.1 Abstracts as Open Access Papers

Results from Table 5-7 point to three main findings. First, as expected, the total number of citations seems to significantly correlate with number of cited open access papers. Second, the negative sign of the coefficients of company size and university collaboration confirm the initial assumptions that smaller companies cite open access papers more than large ones and that they do less so if they collaborate with university.

Table 5-7: Regression analysis results where abstracts are counted as open access papers

Variables	original data only		original data and interpolated data	
	Model I-1	*Model I-2*	*Model I-3*	*Model I-4*
Constant	.3568 (.0015)	.3803 (.3513)	.5453 (.3505)	.6175 (.3441)
All Journals	.01525*** (.0016)	.0152*** (.0016)	.0079*** (.0009)	.0132*** (.0015)
Technological Capability	-.0014 (.0017)	-.0020 (.0014)	.0070 (.0092)	-.0018 (.0016)
Group Membership	.233 (.3205)	.2048 (.3059)	.2179 (.3206)	.2352 (.3156)
Patent Relevance	.1075 (.2491)	.1112 (.2445)	.2818 (.262)	-.0207 (.2402)
University Collaborator	-.9684*** (.1922)	-.9868*** (.1916)	-.0844 (.2024)	-.896*** (.183)
Company Collaborator	.1305 (.2805)	.2074 (.2849)	.0028 (.2121)	.1934 (.2846)
Company Size		-.0006** (.0002)		-.0005* (.0001983)
Number of Observations	825	825	918	918

(robust) standard errors are in parentheses.
*** p<0.001, **p<0.01, *p<0.05, †p<0.1

Table 5-9 helps in the quantification of these effects through the presentation of Incident Rate Ratios (IRR) of different variables. A one standard deviation increase in the number of employees of a given company is associated with a 0.05% decrease in the number of open access papers cited in its patent. Put differently, an increase in the company size by one employee will result in about one (0.9995) less citation to open access papers. Collaborating with university, on the other hand, has a lower impact, as it is associated with a decrease the amount of open access papers cited by 0.4.

The third finding from Table 5-8 is the fact that these effects actually decrease on adding more data (the interpolated data) to the sample. Given that adding more data was shown to make the sample more representative of the population (Table 5-3), it could be possible that these effects will vanish on generalizing the results to the whole population, leading to a rejection of the main hypothesis of this study. However, as noted before, the "population" according to Table 5-3 includes only the companies with entries in the Orbis database. It was shown before that it misses a lot of data and further investigation is needed to determine whether this missing data affects the database's representation of the real population of US pharmaceutical companies.

5.1.3.2 Exclusion of Abstracts

Due to the large proportion of abstracts among open access citations, regression analysis was repeated ignoring citations to abstracts as a way to check the robustness of the results. However, the fact that abstracts are accessible online, free of charge puts them in the same position as open access journals because in both cases it does not cost the companies money to access them. It might be the case that reading abstracts (not full papers) is one mechanism by which companies that cannot afford high journal subscription costs access the scientific information they need. This prevalence of abstracts among the cited literature signifies their importance for industry researchers. On the other hand, usage of abstracts might be alarming in the sense that researchers (because of cost barriers) are forced to settle for less useful sources of information, which hinders the proper diffusion of knowledge. Table 5-8 below reports the correlation coefficient for models where citations to abstracts were dropped.

Table 5-8: Regression analysis results after dropping citations to abstracts from the dataset

Variables	**original data only**		**original data and interpolated data**	
	Model II-1	*Model II-2*	*Model II-3*	*Model II-4*
Constant	.2008 (.3753)	.2158 (.3755)	.5453 (.3505)	.5634 (.3507)
All Journals	.0091*** (.001)	.0090*** (.001)	.0079*** (.0009)	.0078*** (.0009)
Technological Capability	.00524 (.007)	-.0013 (.0013)	.0070 (.0092)	-.0011 (.0014)
Group Membership	.2146 (.3125)	.2036 (.3014)	.2179 (.3206)	.2067 (.3112)
Patent Relevance	.6429** (.2505)	.6570** (.2505)	.2818 (.2622)	.2954 (.2611)
University Collaborator	-.1597 (.1981)	-.1599 (.1966)	-.0844 (.2024)	-.0730 (.2003)
Company Collaborator	-.0271 (.2228)	-.0015 (.2150)	.0028 (.2121)	.0230 (.2060)
Company Size		-.0004† (.0002)		-.0003 (.0002)
Number of Observations	817	817	910	910

(robust) standard errors are in parentheses.
*** p<0.001, **p<0.01, *p<0.05, †p<0.1

As it appears, only the total number of journals and the patent classification are the only two variables with significant coefficients in Table 5-8, Company size is somewhat significant (at the 10% level) but only when using the original data set, not including the interpolated data. This makes it further difficult to confirm the initial assumption that smaller companies use more open access than large ones. However, it is important to note that the sample is somewhat biased against small companies (Table 5-3), which makes rejecting the hypothesis also difficult.

Table 5-9: Incident rate ratios for all explanatory variables (except for year dummies)

Incident Rate Ration (IRR)	Model I-4	Model II-4
All Journals	1.013297	1.007897
Technological Capability	.9982512	.9989467
Group Membership	1.26512	1.229597
Patent Relevance	.9795269	1.343692
University Collaborator	.4082113	.9295952
Company Collaborator	1.213334	1.023312
Company Size	.9995133	.9997117

In general, the very small number of open access citations after excluding abstracts (about 3% of all journal citations both with and without the interpolated data) makes it very difficult to make strong inferences from this regression. In the light of this, there are two important points to be taken into consideration in the interpretation of these results. First, as mentioned before, only articles published in open access journals were counted as open access for the purpose of this analysis, which would naturally result is an underestimation of the counts of open access citations (i.e. because of neglecting green and hybrid OA). This is especially important given that for biomedical research (the object of this study) the existence of a large database like PubMed Central makes available huge amounts of green open access papers.

Second, the issue of how to determine the quality of journals is indeed a very debatable one [128]. However, it can still be argued that the relative short age of all open access journals might put them at a disadvantage with traditional journals that have established good reputation over their long history. Therefore they might not have gained enough popularity (judging by the minority of citations to "born open" journals) among industry researchers to be extensively consulted in their research activities.

5.1.3.3 *Other Findings*

Apart from the regression analysis, one interesting trend found by the study was that patent citations to open access papers have increased over the period of 2005-2013 by an average annual increase of 38%, compared to only 16% for journal citations in general (Figure 5-8a). A similar trend is also observed at the level of patents (Figure 5-8b). The number of patents citing open access journals has sustained and average annual increase of 22% versus only 15% increase in patents citing journals in general. In both graphs, the sharp drop in 2014, is expected due to the very small number of patents granted in 2014 in the dataset (only 16 out of the 825 patents).

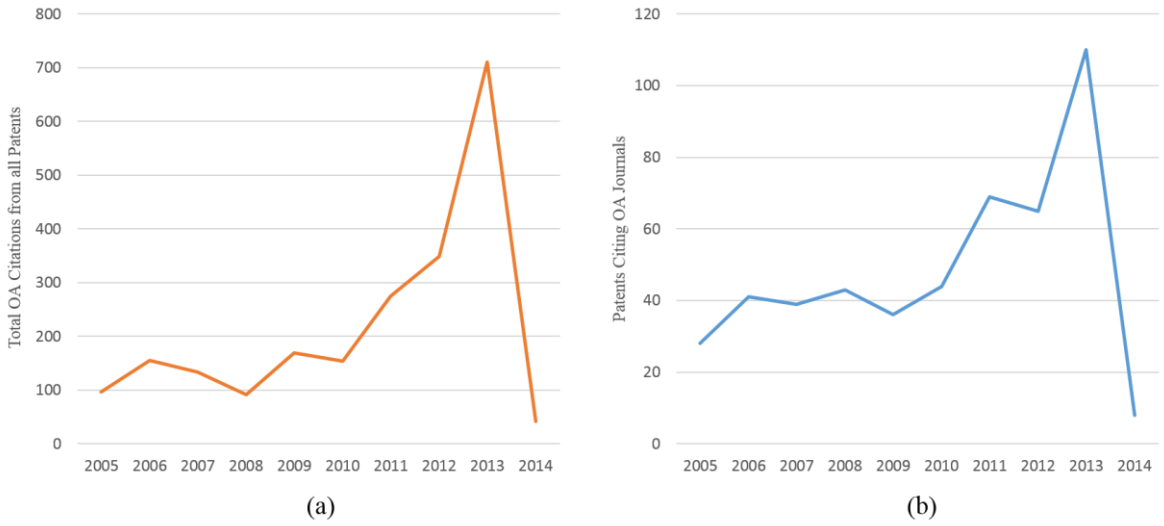


Figure 5-8: Recent fast growth in (a) total citations to open access in all patents and (b) the total number of patents citing open access

5.1.4 Conclusion

This study has introduced a new way to study the impact of open access research in non-academic contexts. Examining citations to open access research is not restricted to patent

documents. The same can be done for other types of documents like clinical guidelines or government reports. The effect of resources available for the citing body (turnover for companies, budget for government research units, etc.) can be a good indicator for their propensity to cite open access journals versus citing journals that require subscription.

Though inconclusive, the results of this study show that open access research might be an efficient way used by small firms to overcome the barrier to knowledge transfer created by high journal subscription costs. This is one type of barrier to knowledge transfer that has not be discussed in the literature before. Failing to reject the initial hypothesis (that smaller companies utilize open access articles more than large companies) does support to some extent the current opinion that many SMEs suffer from high journal prices [41]. They might also support the assumption made by many policymakers (Section 2.3.2) about the impact of open access on innovation.

5.2 Independent Researchers

This section reports an exploratory study aiming to shed light on one of the understudied groups of society that might benefit from the online accessibility of research. Some authors have already discussed this issue but not study was conducted so far (see Section 3.2.7 for more details). It is not clear who these researchers (who publish without stating an academic affiliation) are. There are four main possibilities. First, it is not difficult to imagine individuals who have an interest (and perhaps some previous training) in one area of research and can spare enough time (whether related to or separate from their daily job) to do quality research that can then be published. Another possibility could be when a professional (normally affiliated) researcher is taking a long transition period while moving between two different institutions. It is also possible that the researcher is in fact affiliated with some organization but has reasons why not to mention his/her affiliation on the publication. That is certainly common among researchers who belong to governmental research units who publish results that government might not want to be affiliated with. Consultants are maybe another example of this case. The final possibility is of course researchers who publish from time to time after retirement, especially those who retire early (i.e. women who leave academic positions to spend more time on establishing or maintaining a family).

5.2.1 Methodology

Scopus database was searched (in February 2016) for publications where at least one author states “Independent Researcher” instead of a specific affiliation (e.g. university department

or company). The search returned 844 records. The majority of these publications were coauthored with one or more affiliated researchers. However, a set of 217 publications were solely authored by one (and in a 17 cases multiple) independent researcher(s). Further analysis was conducted to identify the characteristics of this set of publications.

5.2.2 Results and Discussion

Although still insignificant in absolute numbers, the data suggests that the number of publications authored by unaffiliated researchers has been rising over the past decade. A whole decade has separated each two of the first three of these publications (published in 1980, 1990 and 1999, respectively). However, the number of publications by independent researchers has increased from only 3 in 2008 to 50 in 2015 (Figure 5-9). Indeed, part of this trend can be attributed to the general increase in the number of researchers and publications worldwide, but this apparently is not the whole story. On comparing this 1500% increase with only 28% increase in the total number of (Scopus-indexed) publications between 2008 and 2015, one can notice that this trend is difficult to ignore.

In this dataset, almost half of the publications come from the United States (62) and the United Kingdom (38). Australia and India follow by about 11 documents each. The rest are dominated by European and Asian countries along with minimal representation of Africa and Latin America. The majority of publications are journal articles (73%), including review papers. Conference papers come next at 11%, followed by books and book chapters at 8%. Other document types (editorials, letters, research notes, etc.) make up the remaining 8%.

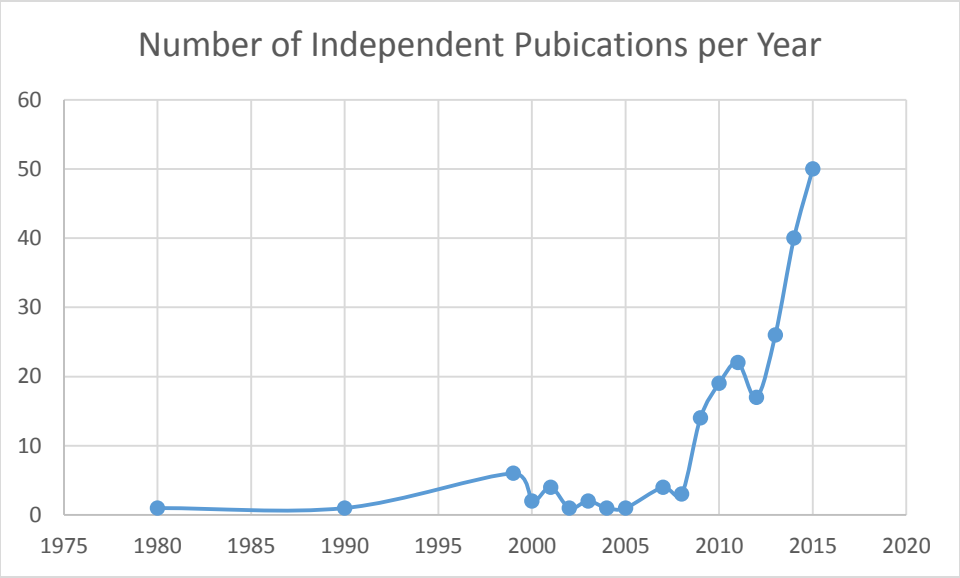


Figure 5-9: Number of publications by "independent researcher" every year

The existence of publications reporting natural science research projects (which might be costly to conduct) called for further investigation into the issue of funding. None of the independently authored papers mentioned funding details. All of the 27 papers that do so are coauthored with affiliated researchers. They fit within the regular expectations of research funding sources (research councils, funding agencies and intramural university funds). It is apparent that some of the papers (solely authored by independent researchers) are based on sophisticated experimental work in both physical and life science fields (Table 5-10), which is generally expensive to conduct and would ideally require significant funding. It is important to note here though that such absence of information does not necessarily mean that no funds were available. It could be that funding sources were not acknowledged on the publications or that some error took place while indexing them.

This issue of the cost of conducting experimental research might be one reason why physical and life science papers solely published by independent researchers are less in proportion compared to those is humanities and social sciences. This assumption can also explain the increase in proportion of life science papers (arguably the most costly) in papers where there are university collaborators (see the right column of Table 5-10: Distribution of papers authored by independent researchers based on research field (using the first classification code provided by Scopus). About 10% of the publications had no classification in Scopus and where categorized to one of the 4 fields above based on the author's judgment.). This assumption, however, does not explain the low representation of papers from arts and humanities, given that research in those fields generally requires much less resources to conduct. The number of arts and humanities journals in Scopus is about one half of that of social science journals, which is not nearly the same ratio of papers in this dataset (in both the middle and the right columns of Table 5-10: Distribution of papers authored by independent researchers based on research field (using the first classification code provided by Scopus). About 10% of the publications had no classification in Scopus and where categorized to one of the 4 fields above based on the author's judgment.). Another assumption to explain this prevalence of papers from social science would be the existence of closely linked professions to many fields of the social science (e.g. economics or public policy). This is unlike the case with many physical or life science disciplines, where not much research takes place outside of universities or credentialed research institutes, making it more complicated for unaffiliated researchers to be involved.

Table 5-10: Distribution of papers authored by independent researchers based on research field (using the first classification code provided by Scopus). About 10% of the publications had no classification in Scopus and where categorized to one of the 4 fields above based on the author's judgment.

Field of Research	Percentage of Papers authored by Independent Researchers alone (n=217)	Percentage of Papers authored by Independent Researchers with collaborators (n=844)
Arts and Humanities	11	5
Social Sciences	49	38
Physical Sciences	17	17
Life Sciences	23	40

Counting the number of references in each publication shows that it is not much different from patterns in scholarly publications in general. Table 5-11 below shows the distribution of references within the dataset. The average publication in this dataset cites about 33 references, which is more than the 20 references cited in all of Scopus publications on average [129]. This raises interesting questions about the ways they access previous literature and the possible benefits open access can have in this regard.

Table 5-11: Citing behavior of Independent Researchers

Number of cited references	Frequency
0-10	64
11-50	104
50-100	39
>100	10
Total	217

Although deeper analysis is required, it appears that the quality of these publications might not be a little less than what is expected of scholarly works in general. Although counting citations is without question a debatable measure of quality, in absence of other measures is can give us some method to compare. Excluding one article that received 484 citations, the 79 articles who have received *any* citations have an average of 4.3 citations per document for

a 16-year citation window. The fact that 138 (64%) of papers were never cited calls for concern given that other studies have demonstrated much lower rates even with shorter citation windows [130].

One limitation for this analysis is that it was based only on those who described their lack of affiliation with the expression “independent researcher”. This might have ignored other ways to describe this status such as “unaffiliated researcher” or “independent scholar”, or even other descriptions that might have been necessary in the context of particular studies (e.g. “concerned citizen” or “environmental activist”).

5.2.3 Further Research

Further research into this topic can inform the debate about the (de)institutionalization of research practice, as well as the impact of the Open Access movement on the traditional gatekeeping function of journals in academia. First, we need to identify the real status of those who publish as “Independent Researchers”. In other words, it is important to know the relative proportions of amateur researchers, researchers in transition, or those who have reasons not to mention their affiliations or have already retired. Knowing these proportions will help us specify why this trend is rising and might save researchers the effort of further investigations if it turns out that the majority of this “unaffiliated researchers” are based in places that offer them large-scale journal access (e.g. universities or large corporations).

Second, more insights are also needed into how these researchers get funding for their studies. This can be extended to include other potential challenges too (e.g. access to previous

literature, use of research equipment, rejection for lack of affiliation, etc.) in addition to their strategies to overcome those challenges. This is especially important to investigate in the context of papers that belong to physical and life sciences (about half of the studies in the used sample).

Third, analyzing the sources these independent researchers cite in their publications can provide pertinent insights on their relationship with open access. This relationship (along with the two issue raised above) can also be investigated through an online survey of these researchers using the correspondence emails they provide on publication.

6 Conclusion & Policy Implications

This concluding chapter is organized in three sections. The first one summarizes the findings of this work as a whole and derives conclusions from its different sections (in terms of answers to research questions). The second part suggests some directions for future research, while the third part presents the possible implications these findings can have for policymakers involved in the open access debate (e.g. in ministries of science and technology or in research funding agencies).

6.1 Answers to Research Questions

RQ1: Which groups of people outside the academic community benefit (or would potentially benefit) from open access to scholarly research?

Those from outside the academic community who use research papers need them for three reasons: to do research, to enhance their practice or for personal use. Further breakdown is provided in Figure 4-1. What is important to note here is to what we know about each of these categories.

Citizen scientists, unaffiliated researchers and those who use research papers for personal uses all need to be better understood. Available evidence suggests that they are likely (albeit with different degrees) to benefit from open access research. The problem with these groups is that the contexts in which the encounter research is in itself not clear, let alone how they deal with issues of access. For example, it is not clear if the importance of open access research to patients depends on their capacity to do research on their own versus on them

taking part in a patient group that identifies a more clear research agenda. The study pf papers published by “independent researchers” (Section 5.15.1) has provided a first step in this direction by showing that the documents these groups publish can be a resource for studying their usage of academic research. It was shown that independent researchers come from different fields, use references at a level comparable to affiliated researcher and potentially produce work of similar quality.

Because evidence-based policymaking is an already mature field of research, we have a somewhat good understanding of the mechanism by which policymakers engage with academic research. However, understanding the extent to which open access research in particular can contribute to better policymaking still needs much more research that what is currently available. This contention also applies to workers in the nonprofit sector given the similarity of the role played by academic research in their work to that in government policymaking.

Among professionals whose job is closely tied with academic research, previous studies have addressed two groups, although unequally. Some signals point to the value open access research might have in legal practice. However, medical practitioners are studied more extensively. Some evidence shows that they might actually benefit from open access research. It is not clear though whether this is homogenous across different types of medical practitioner (e.g. nurses in general hospitals, family doctors, dentists in private practice). Another issue that also needs deeper analysis is the extent to which they benefit from research papers *directly* as opposed to other secondary sources based on them like clinical guidelines

or clinical decision support services (e.g. UpToDate). In all cases, the huge body of literature available on what is called “evidence-based practice” should always be used to contextualize any further studies in this regard.

RQ2: How is the company size related to its usage of open access research?

Regarding industrial R&D, we know that research is important for many industrial endeavors and the process of knowledge diffusion from academic research to innovative applications is an extensively studied topic. In terms of the impact of open access, most of the available evidence (as shown in Section 3.2.2) seems to focus on SMEs and suggests that SMEs suffer from the high prices of journal subscriptions. However, this is mainly according to what is claimed by SMEs themselves. This study was the first to use a methodology other than surveys to investigate this claim. Because of the large amount of missing company data and the consideration of only gold open access (leaving out green and hybrid), this study was not conclusive regarding the relationship company size to usage of open access. Some signals point to the possibility initially predicted (that smaller companies use open access more) but further research is needed using more data. This is still an important insight into the issue given that the high journals subscription cost is a new type of barrier to the process of knowledge transfer that has been overlooked in the currently available literature.

RQ3: How can we study the usage of open access by unaffiliated researchers?

Using bibliometric techniques to study the usage of open access literature by independent researchers was shown to be a good way to understand this issue. This study also represented

the first attempt to systematically approach the topic of unaffiliated researchers. However, due to the inability to identify whether those researchers are actually unaffiliated (versus being affiliated but with reasons to hide their affiliation), it was important to suggest further research using survey methods before further analysis of open access usage is conducted.

6.2 Future Research

The impact of open access research outside the academic community is a complex and versatile issue. It is important to approach it with creative approaches given the difficulty in quantifying any of these different impacts. All of the studies cited here (as examples of usage of research outside academia) come from a variety of fields. Therefore, any future research efforts should take this into account. A research program would ideally consider the different examples noted in the framework proposed by this study as well as try to add new groups of users. For each group (SMEs, NPOs, patient organizations, parents, etc.) one or two indicators should be developed. While an indicator for one group of users can inform the development of that of a different group, attempts to generalize the use of the same indicator for many groups will probably fail to represent reality. In essence, that would be similar to using the number of citations as a measure of quality for journals (i.e. impact factor), which proved to be inadequate in a number of fields. It is imperative to accept that the definition of "impact" varies depending on context and that one-size-fits-all indicators compromise the quality of any analysis [23].

6.3 Policy Implications

Identifying who the beneficiaries are is very essential to inform any debate a global transition to open access. Apart from partly supporting many policymakers' assumption that mandating open access would impact SMEs and consequently innovation (Section 2.3.2), the important issue raised by this study is that the Open Access movement have much more stakeholders than those currently involved in the debate. This is important in two ways. First, the debate can be improved to a great extent by introducing the views and the concerns of other groups of people who are affected by the mode of scholarly communication. Second, regardless of the extent to which excessive publisher profits contribute to the problem, the controversy on who should fund "fair" open access can be simplified if more beneficiaries are involved, especially those who extensively consume academic literature and contribute very little to it (e.g. industry researchers). There is no reason why industries that bear 15-17% of subscription costs [38] be excluded from contributing to the cost of open access provision. Some leading publishers even view this segment of open access users as an expanding market for their journal subscription schemes [38]. It has already been suggested that industry associations (e.g. the Biotechnology Industry Organization) manage schemes of group subscriptions to remedy the problem of SME access to research [41]. It is not difficult to imagine similar scenarios taking place in an open access world.

A similar argument can be made for some government agencies that already subscribe to journals. A good example already exists whereby one government organization (with a stake in the accessibility of research) can directly support open access. A mutual agreement with

the European Observatory on Health Systems and Policies (an intergovernmental organization) made it possible for the journal *Health Policy* to provide free access to selected policy-relevant articles for the benefit of everyone [131]. Indeed, the more these collaborations take place, the more seamless the transition to open access can be. Another transition mechanism can be achieved through the involvement of research intensive businesses in agreements of national site licenses that many governments now negotiate.

In conclusion, the issue of societal benefits of open access can prove to be very complex and manifold. However, approaching it in the way suggested by this study can take the debate on access to research to a whole new level, by reframing it as a social issue, rather than one that is just relevant to the academic community.

7 References

1. Price DJdS. Little science, big science New York: Columbia University; 1963.

2. Morrison HG. Freedom for scholarship in the internet age. PhD Thesis. Simopn Fraser University, School of Communication; 2012.

3. Dingley B. U. S. Periodical Prices. American Library Association, Association for Library Collections & Technical Services; 2005.

4. UK Office of Fair Trading. The market for scientific, technical and medical journals. OFT396. ; 2002.

5. Mark Ware Consulting. Access by UK small and medium-sized enterprises to professional and academic information. commissioned by Publishing Research Consortium. ; 2009.

6. Van Noorden R. Open access: The true cost of science publishing. Nature (News Feature). 2013 March; 495(7442).

7. Suber P. Open access Cambridge: MIT Press; 2012.

8. Willinsky J. The Nine Flavours of Open Access Scholarly Publishing. Journal of Postgraduate Medicine. 2003 July; 49(3).

9. Prosser D. The costs of double dipping. [Online].; 2015 [cited 2016 June. Available from: http://www.rluk.ac.uk/about-us/blog/the-costs-of-double-dipping/.

10. Suber P. The taxpayer argument for open access. SPARC Open Access Newsletter. 3003 September.

11. Tennant JP, Waldner F, Jacques DC, Masuzzo P, Collister LB, Hartgerink CHJ. The academic, economic and societal impacts of Open Access: an evidence-based review. F1000Research. 2016 April; 5(632).

12. Swan A. The Open Access citation advantage: Studies and results to date. Project Report. s.n.:; 2010.

13. Hitchcock S. The effect of open access and downloads ('hits') on citation impact: a bibliography of studies. [Online].; 2013 [cited 2016 October. Available from: http://opcit.eprints.org/oacitation-biblio.html.

14. SPARC-Europe. The Open Access Citation Advantage Service (OACA). [Online].; 2016 [cited 2016 October. Available from: http://sparceurope.org/what-we-do/open-access/sparc-europe-open-access-resources/open-access-citation-advantage-service-oaca/.

15. McCabe MJ, Snyder CM. Open Access as a Crude Solution to a Hold-up Problem in the Two-Sided Market for Academic Journals. NBER Working Paper Series. 2016 May;(No. 22220).

16. University of California Libraries. Pay It Forward: Investigating a Sustainable Model of Open Access Article Processing Charges. Mellon Foundation; 2016.

17. Jiménez AC, Willinsky J, Boyer D, Col GD, Golub A. Why an open access publishing cooperative can work - A proposal for the AAA's journal portfolio. Journal of Ethnographic Theory. 2015; 5(2).

18. Zuccala A. Open access and civic scientific information. Information Research. 2010 March; 15(1).

19. Davis PM, Walters WH. The impact of free access to the scientific literature: a review of recent literature. Journal of the Medical Library Association. 2011; 99(3).

20. Pinfield S. Making open access work: The 'state-of-the-art' in providing open access to scholarly literature. Online Information Review. 2015; 39(5): p. 604-636.

21. Bankier JG, Chatterji P. 100 Stories: The Impact of Open Access. bepress; 2016.

22. Research Information Network (RIN). Monitoring Progress in the Transition to Open Access. ; 2014.

23. Bornmann L. Measuring the societal impact of research. EMBO Reports. 2012; 13(8).

24. Martin BR, Tang P. The benefits from publicly funded research. SPRU Electronic Working Paper Series. 2007 June.

25. Sarewitz D. Does Science Policy Matter? Issues in Science and Technology. 2007 Summer; XXIII(4).

26. Meijer I, Calero-Medina CM, Abma-Schouten R. Societal Use of Research (SURe): stakeholder perceptions and data method development. In Proceedings of 18th International Conference on Science and Technology Indicators; 2013. p. 237-241.

27. Mostert S, Ellenbroek SP, Meijer I, van Ark G, Klasen EC. Societal output and use of research performed by health research groups. Health Research and Policy Systems. 2010; 8(30).

28. Godin B, Dore C. Measuring the Impacts of Science; Beyond the Economic Dimension. [Online].; 2005 [cited 2017 March. Available from: http://www.csiic.ca/PDF/Godin_Dore_Impacts.pdf.

29. Zuccala A. Chapter 8 - The Layperson and Open Access. Annual Review of Information Science and Technology. 2009.

30. Callon M. The Role of Lay People in the Production and Dissemination of Scientific Knowledge. Science, Technology and Society. 1999 March; 4(1).

31. Björk BC. A model of scientific communication as a global distributed information system. Information Research. 2007 January; 12(2).

32. Houghton J, Rasmussen B, Sheehan P, Oppenheim C, Morris A, Creaser C, et al. Economic Implications of Alternative Scholarly Publishing Models - Exploring the costs and benefits. commissioned by JISC. ; 2009.

33. Wahab SA, Rose RC, Uli J, Abdullah H. A Review on the Technology Transfer Models, Knowledge-Based and Organizational Learning Models on Technology Transfer. European Journal of Social Sciences. 2009; 10(4).

34. Eckl VC. Creating an Interactive-Recursive Model of Knowledge Transfer. In DRUID Conference; 2012; Copenhagen.

35. Sung EK, Gibson DV. Knowledge and technology transfer: levels and key factors. IC2 Publications. University of Texas at Austin, IC2 Institute; 2000.

36. Bozeman B. Technology transfer and public policy: a review of research and. Research Policy. 2000; 29.

37. Eckl VC. Barriers of knowledge transfer. In DRUID Conference; 2012; Copenhagen.

38. Ware M, Mabe M. The STM Report (Fourth Edition). The Hague:; 2015.

39. Content Mine. [Online]. Available from: www.contentmine.org.

40. Cohen WM, Levinthal DA. Absorptive Capacity: A New Perspective on Learning and Innovation. Administrative Science Quarterly. 1990; 35(1): p. 128-152.

41. Lyman S. Industry access to the literature (letter to the editor). Nature Biotechnology. 2011 July; 29(7): p. 571-572.

42. CIBER. Access to scholarly content: gaps and barriers. commissioned by RIN, JISC & PRC. ; 2011.

43. Houghton J, Swan A, Brown S. Access to Research and Technical Information in Denmark. commissioned by the Danish Agency for STI and Denmark's Electronic Research Library. ; 2011.

44. Cohen WM, Nelson RR, Walsh JP. Links and Impacts: The Influence of Public Research on Industrial R&D. Management Science. 2002; 48(1).

45. Irwin H, More E. Technology transfer and communication: lessons from Silicon Valley, Route 128, Carolina's Research Triangle and hi-tech Texas. Journal of Information Science. 1991; 17(5).

46. Sun PYT, Scott JL. An investigation of barriers to knowledge transfer. Journal of Knowledge Management. 2005; 9(2).

47. Disterer G. Individual and Social Barriers to Knowledge Transfer. In Proceedings of the 34th Hawaii International Conference on System Sciences; 2001.

48. Registry of Open Access Repository Mandates and Policies. [Online].; 2016 [cited February. Available from: http://roarmap.eprints.org/.

49. Prosser D. Public Policy and the Politics of Open Access. LIBER QUARTERLY. 2007; 17(2).

50. Prosser DC. Current (European) Developments in Scholarly Communication. Liber Quarterly. 2008; 18(3/4).

51. Kramer B, Bosman J. Innovations in scholarly communication - global survey on research tool usage. F1000 Research. 2016 April; 5(692).

52. Group RW. Monitoring Progress in the Transition to Open Access. ; 2014.

53. ElSabry E. Mapping Open Access Societal Impact. Research Ideas and Outcomes. 2017; 3(e11743).

54. Open Access Directory. Declarations in support of OA. [Online].; 2016 [cited 2017 3 10. Available from: http://oad.simmons.edu/oadwiki/Declarations_in_support_of_OA.

55. Flis I, Haslbeckband J, Noone C. European Student Scientifc Literature Access Study..

56. Vincent-Lamarre P, Boivin J, Gargouri Y, Larivière V, Harnad S. Estimating open access mandate effectiveness: The MELIBEA score. Journal of the Association for Information Science and Technology. 2016; 67(11): p. 2815-2828.

57. Teplitzky S, Phillips M. Evaluating the Impact of Open Access at Berkeley: Results from the 2015 Survey of Berkeley Research Impact Initiative (BRII) Funding Recipients. College & Research Libraries. 2016; 77(5): p. 568-581.

58. Tsoukala V, Angelaki M. Open Access Policy Guidelines and Template for Funders. Policy Guidelines. ; 2015.

59. Picarra M. Open Access to scientific information: facilitating knowledge transfer and technological innovation from the academic to the private sector. Briefing Paper. ; 2015.

60. Swan A. Policy Guidelines for the Development and promotion of Open Access. Paris:; 2012.

61. Waltham M. Open access - the impact of legislative developments. Learned Publishing. 2005; 18: p. 101-114.

62. Houghton J. Costs and Benefits of Alternative Scholarly Publishing Models: Lessons and Developments. In Publishing in the networked world :transforming the nature of communication : 14th International Conference on Electronic Publishing; 2010; Helsinki.

63. EU Framework Program for Research and Innovation. Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020 (Version 1.0). ; 2013.

64. Holdren JP. Increasing Access to the Results of Federally Funded Scientific Research. ; 2013.

65. Brossard D, Lewenstein B, Bonney R. Scientific knowledge and attitude change: The impact of a citizen science project. International Journal of Science Education. 2005 July; 27(9): p. 1099–1121.

66. Bailey CW. Open Access Bibliography. [Online].; 2005. Available from: http://digital-scholarship.org/oab/oab.htm.

67. Bailey CW. Transforming Scholarly Publishing through Open Access: A Bibliography. [Online].; 2010 [cited 2015 December. Available from: http://digital-scholarship.org/tsp/transforming.htm.

68. Chung J, Tsay MY. A Bibliometric Analysis of the Literature on Open Access in Scopus. Qualitative and Quantitative Methods in Libraries. 2015; 4.

69. Duzyol G, Taskin Z, Tonta Y. Mapping the Intellectual Structure of the Open Access Field Through Co-citation Analysis. In IFLA Satellite Pre-conference: Open Access to Science Information Trends, Models and Strategies for Libraries, Crete; 2010; Crete.

70. Grandbois J, Beheshti J. A bibliometric study of scholarly articles published by library and information science authors about open access. Information Research. 2014 December; 19(4).

71. Pablo-Alperin J. The public impact of Latin America's approach to open access. PhD Thesis. Stanford University, Stanford Digital Repository; 2015.

72. Sato S, Yuki S, Immura H. Citizens Demand for OA to Academic Papers. In Japan Society of Library and Information Science Spring Workshop; 2011.

73. Krane D. Large Majorities of U.S. Adults Support Easy – and Free – Online Access to Federally-Funded Research Findings on Health Issues and Other Topics. Poll Report. Rochester, NY: Harris Interactive; 2006.

74. Fox S, Duggan M. Health Online 2013. Internet & American Life Project. Pew Research Center; 2013.

75. Smith SR. Open Access Journals and Public Engagement. PS: Political Science & Politics. 2015 September; 48(S1).

76. Willinsky J. Derridas right to philosophy then and now. Educational Theory. 2009 August; 59(3).

77. Sato S, Itsumura H. How Do People Use Open Access Papers in Non-academic Activities? : A Link Analysis of Papers Deposited in Institutional Repositories. Library, Information and Media Studies. 2011; 9(1).

78. Willinsky J. What open access research can do for Wikipedia. First Monday. 2007 March; 12(3-5).

79. Teplitskiy M, Lu G, Duede E. Amplifying the Impact of Open Access: Wikipedia and the Diffusion of Science. Journal of the Association of Information Science & Technology. 2016 October; early view.

80. Parsons D, Willis D, Holland J. Benefits to the Private Sector of Open Access to Higher Education and Scholarly Research. UK Open Access Implementation Group; 2011.

81. Abe K, Yaguchi M, Shimase S. A follow-up survey on e-journal usage among companies and the academia in Japan. Journal of Information Processing and Management. 2011 October; 54(7).

82. DeepDyve. The 2016 DeepDyve Report. ; 2016.

83. Ozcan Y, Bryan KA. The Impact of Open Access Mandates on Invention. unpublished. 2016 June.

84. Willinsky J. Policymakers' Online Use of Academic Research. Education Policy Analysis Archives. 2003 January; 11(2).

85. Nwagwu WE, Iheanetu O. Use of Scientific Information Sources by PolicyMakers in the Science and Technology Sector of Nigeria. African Journal of Library, Archives and Information Science. 2011; 21(1).

86. Willinsky J. As Open Access is Public Access, Can Journals Help Policymakers Read Research? Canadian Journal of Communication. 2004; 29.

87. Look H, Marsh K. Benefits of Open Access to Scholarly Research to the Public Sector. UK Open Access Implementation Group; 2012.

88. Rich RF. Social Science Information and Public Policy Making San Francisco: Jossey-Bass; 1981.

89. Beddoes D, Brodie E, Clarke R, Sin DCH. Benefits of Open Access to Scholarly Research for Voluntary and Charitable Sector Organisations. UK Open Access Implementation Group; 2012.

90. Moorhead LL, Holzmeyer C, Maggio LA, Steinberg RM, Willinsky J. In an Age of Open Access to Research Policies: Physician and Public Health NGO Staff Research Use and Policy Awareness. PLOS One. 2015 July; 10(7/e0129708).

91. Davies K. The information-seeking behaviour of doctors: a review of the evidence. Health Information and Libraries Journal. 2007; 24.

92. Hardisty DJ, Haaga DAF. Diffusion of Treatment Research: Does Open Access Matter? Journal of Clinical Psychology. 2008; 64(7).

93. Bryant N. Navigating Access to Addictions Literature: The Hubert H. Humphrey Fellowship Experience. In el-Guebaly N, Carrà G, Galanter M, editors. Textbook of Addiction Treatment: International Perspectives.: Springer-Verlag Italia; 2015. p. 2439-2450.

94. Spedding S. Open Access Publishing of Health Research: Does Open Access Publishing Facilitate the Translation of Research into Health Policy and Practice? Publications. 2016 January; 4(2).

95. O'Keeffe J, Willinsky J, Maggio L. Public Access and Use of Health Research:An Exploratory Study of the National Institutes of Health (NIH) Public Access Policy Using Interviews and Surveys of Health Personnel. Journal of Medical Internet Research. 2011; 13(4/e97).

96. Lawton A, Flynn E. The Value of Open Access Publishing to Health and Social Care Professionals in Ireland. Ariadne. 2015; 73.

97. Andrews JE, Pearce KA, Ireson C, Love MM. Information-seeking behaviours of practitioners in a primary care practice-based research network (PBRN). Journal of the Medical Library Association. 2005; 93.

98. Chan L, Arunachalam S, Kirsop B. The chain of communication in health science: from researcher to health worker through open access. Open Medicine. 2009; 3(3).

99. Maggio LA, Moorhead LL, Willinsky JM. Qualitative study of physicians' varied uses of biomedical research in the USA. BMJ Open. 2016 September; 6(e012846).

100. Maggio LA, Steinberg RM, Moorhead L, O'Brien B, Willinsky J. Access of primary and secondary literature by health personnel in an academic health center: implications for open access. Journal of the Medical Library Association. 2013 July; 101(3).

101. Maggio L, Alperin JP, Moorhead L, Willinsky J. Can Your Doctor See the Cancer Research Reported in the News? Can you? [Online].; 2017 [cited 2017 February 22.

Available from: https://medium.com/@lauren.maggio01/can-your-doctor-see-the-cancer-research-reported-in-the-news-can-you-beb9270c301f.

102. Danner RA. Open Access to Legal Scholarship: Dropping the Barriers to Discourse and Dialogue. Journal of International Commercial Law and Technology. 2012; 7(1).

103. Schwartz DL, Petherbridge L. The Use of Legal Scholarship by the Federal Courts of Appeals: An Empirical Study. Cornell Law Review. 2010; 96(6).

104. Scherlen A, Robinson M. Open Access to Criminal Justice Scholarship. Journal of Criminal Justice Education. 2008 March; 19(1).

105. Danner D, Popa R, Palfrey J, Germain C, George P, McMasters J, et al. Durham Statement on Open Access to Legal Scholarship. [Online].; 2008 [cited 2016 Novemebr. Available from: https://cyber.harvard.edu/publications/durhamstatement.

106. Science Commons. Open Access Law Program. [Online]. [cited 2017 April. Available from: http://www.sciencecommons.org/projects/publishing/oalaw/.

107. Terry S. In the public interest. C&RL News. 2005 July/August.

108. Terry SF. Disease advocacy organoizations catalyze translational research. Frontiers in Genetics. 2013 June; 4(101).

109. Alliance for Taxpayer Access. [Online].; 2004 [cited 2017 March. Available from: http://www.taxpayeraccess.org/.

110. Line M. Access to documents by the independtent researcher. Interlending & Document Supply. 2011; 29(4).

111. Roff S. The return of the armchair scholar. Journal of Scholarly Publishing. 2005 January; 36(2).

112. Lave R. Neoliberalism and the Production of Environmental Knowledge. Environment and Society: Advances in Research. 2012; 3.

113. Kullenberg C, Kasperowski D. What Is Citizen Science? – A Scientometric Meta-Analysis. PLOS One. 2016 January; 11(1).

114. Kimura AH, Kinchy A. Citizen Science: Probing the Virtues and Contexts of Participatory Research. Engaging Science, Technology, and Society. 2016; 2.

115. Davis PM. Open access, readership, citations: a randomized controlled trial of scientific journal publishing. The FASEB Journal. 2011 March; 25(7).

116. Bauer MW. The Evolution of Public Understanding of Science—Discourse and Comparative Evidence. Science, Technology & Society. 2009; 14(2).

117. Zuccala A, Verleysen F, Engels T. The societal impact of history texts - Lay Readership and the 'Altmetric' value of Goodreads. In 19th Nordic Workshop on Bibliometrics and Research Policy; 2014; Reykjavik.

118. Gibbons M, Limoges C, Schwartzman S, Scott P, Trow M. The New Production of Knowledge: The Dynamics of Science and Research in Contemporary Societies London: Sage Publications Ltd.; 2007.

119. Jaffe AB, Trajtenberg M, editors. Patents, Citations & Innovations: a Window on the Knowledge Economy. 1st ed. Cambridge: MIT Press; 2002.

120. Kuznets S. Inventive Activity: Problems of Definition and Measurement. In NBER , editor. The Rate and Direction of Inventive Activity: Economic and Social Factors. New Jersey: Princeton University Press; 1962. p. 19 - 52.

121. Griliches Z. Patent Statistics as Economic Indicators. In Griliches Z, editor. R&D, Patents and Productivity.: University of Chicago Press; 1984.

122. Narin F, Hamilton KS, Olivastro D. The increasing linkage between U.S. technology and public science. Research Policy. 1997; 26.

123. ElSabry E, Sumikura K. Who needs access to researcch? the case of pharmaceutical industry. In 11th Munin Conference on Scholarly Publishing; 2016; Tromso.

124. Cotropia CA, Mark A. Lemley BS. Do applicant patent citations matter? Research Policy. 2013 February; 42.

125. Fedoryszak M, ŁB, Tkaczyk D, Wojciechowski K. Methodology for Evaluating Citation Parsing and Matching. In Bembenik R, Skonieczny L, Rybinski H,

Kryszkiewicz M, Niezgodka M, editors. Intelligent Tools for Building a Scientific Information Platform. Berlin: Springer; 2013. p. 145-154.

126. OECD. Frascati Manual: Guidelines for collecting and reporting data on research and experimental development. Paris: OECD Publishing; 2015.

127. Directory of Open Access Journals. [Online]. [cited 2016 February. Available from: www.doaj.org.

128. Brembs B, Button K, Munafò M. Deep impact: unintended consequences of journal rank. Frontiers in Human Neuroscience. 2013 June;(00291).

129. Scopus. Content Coverage Guide. [Online].; 2016 [cited 2017 April. Available from: https://www.elsevier.com/solutions/scopus/content.

130. Larivière V, Gingras Y, Archambault É. The decline in the concentration of citations, 1900–2007. Journal of the Association for Information Science and Technology. 2009 April; 60(4).

131. Busse R. Collaboration between Health Policy and the European Observatory on Health Systems and Policies enables Open Access to selectedpolicy-relevant articles. Health Policy. 2013; 111.