PEER Economics Report

Centro ASK, Università Bocconi

November 2011

Università Commerciale
Luigi Bocconi

ASK
Art, Science
and Knowledge

Table of content

Executive summary

This study considers the effect of large-scale deposit on scholarly research publication and dissemination (sharing of research outputs), beginning with the analysis of publishers and institutions managing repositories and their sustainability. The study associates costs with specific activities, performed by key actors involved in research registration, certification, dissemination and digital management: authors, the scholarly community, editors, publishers, libraries, readers and funding agencies. Contrary to most of the existing literature, the study analyses cost structures of individual organizations. The focus of this study is therefore to provide context for the costs to specific organizations and to their choices in terms of scale and scope.

The current competitive scene shows the presence of traditional academic journals, subscription based (SB) together with a growing number of Open Access (OA) digital-only journals. Both types of journals may be published by commercial publishers, university presses, learned societies and not-for-profit organizations. Publication in an OA journal is an alternative to publication in a traditional SB academic journal and the two models are often in direct competition. In terms of business model, journals differ with respect to revenue drivers and financial leverage opportunities based on copyright. Publication in SB journals is typically free for the author; costs of published articles (including peer review, editorial costs, marketing, and all other costs) are covered via subscriptions and pay per view paid for by libraries and by individual users.

The OA movement was started by groups of researchers who wanted to maximise the visibility of publicly funded research, as well as to counteract the increases in subscription rates (ARL Statistics 2007-2008 p. 11), and has been backed by funding agencies and libraries. The argument for public intervention into an imperfect market was related to the opportunity to make publicly funded research more equitably accessible to all stakeholders worldwide. The emergence of several high reputation OA journals, together with an array of journals started and managed with limited resources is an indication of the success of policies aimed at the opening of the competitive space. The presence of a growing number of OA journals has contributed to an increase in competitiveness in the scholarly publishing industry in attracting submitted papers, readers' attention and resources.

In parallel with OA journals, repositories, particularly subject based, emerged as opportunities for authors to increase effectiveness and efficiency of their reputation building by giving their research early visibility and by allowing researchers' communities to cooperate. Moreover, repositories were started to give visibility to institutional or disciplinary research outcomes carried out in specific institutions or financed by particular funding agencies.

Today, the distinctions between the three models (SB, OA or repositories) are blurred, although it is becoming clear that the success of OA journals and repositories – as is the case for SB journals – depends on the strategies of individual players, and not merely OA status. The success of BioMed Central and PLoS proves that OA status does not equate in principle to lower quality of research as was suggested initially by some concerned authors. At the same time, OA status does not in itself automatically lead to higher citation and visibility for

the authors. In the case of repositories, while some (such as REPEC and ArXiv) succeeded in becoming a starting point and not just a destination in scholarly search (i.e. a site actively searched for and not referred to via a keyword based search), many other repositories are less visible.

This study analyses 22 organizations involved with journal article publication and dissemination. Data were gathered via literature and public document analysis, as well as through individual in-depth interviews in order to assess the cost structure of publishers, OA journal publishers and institutions managing repositories and the conditions for their sustainability. Results of the empirical research on the costs associated with research certification, publication and digital management by a sample of journal publishers and repositories highlights the following elements:

- The average cost of content certification per article published for the publishers considered is around 250 USD; the cost includes only salary costs and external fees paid for organizing and managing peer review. No indication of significant economies of scale may be traced at the editorial level, except for submission tracking.
The incidence of content certification on the total cost of archived articles varies significantly among publishers in the sample, and it depends on the journal rejection rate as well as the complexity and length of the review process. To reduce the impact of costs of content certification on total costs, publishers include in their portfolio journals with differing rejection rates.

- The average cost of publishing (including metadata) ranges from 170 to over 400 USD per article, and is influenced by make/buy decisions and by journal size; publishers publishing journals in English are favoured in outsourcing to low-wage countries.

- In order to make content accessible, it has to be managed via a digital platform allowing content management, storage and accessibility. Costs associated with digital platforms vary significantly, depending on whether the platform is proprietary or based on open source (OS) software, on the age and characteristics of the platform, on the number of articles and documents stored and on the complexity of the platform in terms of services offered to readers and authors. Therefore, the incidence of platform investment on article costs is hard to calculate and shows a high variance across the publishers interviewed. Moreover, in the case of some repositories and OA publishers, platform set up investments have been covered by a grant, so that cost structure does not include initial investment repayments. OS platforms allow for shared costs of upgrades and easier interoperability, but the research team was unable to quantify cost differentials with proprietary platforms in this respect. Maintenance costs are somewhat easier to account for and discrepancies among companies in the sample are much lower. Publishers interviewed report yearly maintenance costs ranging from USD 170k to 400k. Incidence of these costs on average cost per article depends on the size of the publisher in terms of number of journals published, number of articles per journal, and the complexity of the platform in terms of services offered to authors and users.

- The availability of OS publishing platforms allows for a drastic reduction of certification, publication and platform management costs, thus reducing barriers to the self-organization of groups of researchers to publish journals, and favouring the start up of new journals. It has to be noted though that a significant portion of these costs is transferred researchers themselves, who provide voluntary labour associated not only with content certification, but also with all

aspects related to management of the journal. Given the low scale effect on certification and publication costs, it is possible that new OA journals relying heavily on volunteer work publish a very limited number of papers.

- Management of repositories is cost effective from an operational point of view and in line with the mandatory functions of the institutions analysed. The use of OS software and the effort required to participate in collaborative projects allow for limited direct costs. However, the impact on making scholarly research openly accessible is related to the availability of a critical mass of well-organized, visible and easily accessible scholarly research. Additionally, the limited resources devoted to repository management make it difficult to enhance services, thus affecting repository effectiveness detrimentally.

- Moreover, many repositories are characterised by a high level of sunk costs. For instance, in the case of organizations coordinating several research institutions, costs of uploading articles and adding metadata are transferred to participating libraries; the costs of software maintenance and upgrades are hard to identify, as they are sunk in overall information technology (IT) budgets. The ability of the organizational unit in charge to enforce the necessary mandates are limited and often related to time-consuming internal diplomacy, which is not accounted for.

The current state of scholarly publication and communication requires that, in order to be sustainable, business models must simultaneously address the following six issues:

- The growth in the supply of documents, as a result of at least four factors: globalization of the research community, growth of published output, public availability of content at different stages of development, and the availability of documents such as reports and newsletters which are published but not recognized as academic publications. All these materials require energy and resources to be selected, verified and organized; the set up of platforms hosting new types of output requires additional ad hoc investments.

- The fragmentation of the industry and interdependence among players: Besides the development of new models for content publication and distribution, an array of start-ups is emerging to deal with specific aspects of scholarly research publishing, dissemination and preservation. This group of "cultural entrepreneurs" (partially private, partially not-for-profit), once established, is likely to increasingly contribute to the transformation of the industry towards a higher importance attributed to information services in addition to certified content.

- The incentives set up by a variety of research institutions: Funding agencies are increasingly putting pressure on scholars to make their research publicly available, thus favouring de facto OA journals; at the same time, incentives to researchers by research institutions to publish on highly visible journals may hinder the diffusion of OA journals, making the scenario of a substitution of SB to OA unlikely[1]. The interplay of incentives of different nature affects OA

[1] On the benefits of a more aggressive mandatory policy see for instance Houghton et al. 2010. Commenting abut the sustainability of OA journals, Suber (2009a) acknowledges the different level of maturity of different business models: "if comparatively little is spent today on OA journals, that says more about the history of

diffusion across disciplines, while in general putting pressure on smaller and less visible journals, both SB and OA.

- The economy of attention: In spite of increased effectiveness in making content available and organized so as to increase the options for researchers to access the information they need and to increase quality and efficiency of research production, the number of articles a researcher will read and cite is limited and competition for researcher's attention is fierce. Therefore, the value of a publication will be increasingly related to the amount and quality of information services provided to the research community; both repositories and journals will have to develop services to authors and readers, undoubtedly with benefits to the scientific community at large, but with a pressure on their costs and critical mass.

- The preservation of memory: As the amount and variety of content produced increase, the problem of guaranteeing appropriate preservation of published research and of other relevant material in digital form becomes increasingly important. Specific resources need to be devoted in order to provide effective, efficient and secure storage and accessibility of content over time.

- The overall financial crisis and a generalised strain on resources: The current economic crisis faced by European countries will put a pressure on available resources at all levels, increasing the competition among research groups to access resources and the competition between journals to attract research outputs. At the same time, cooperation among players is also possible, to address simultaneously the needs of authors and readers. It is likely that funding agencies will pay more attention to costs and benefits associated with alternative resource allocation, and pressure on journals and repositories will increase to specifically address sustainability.

Given these issues, OA and SB journals will have to become more active in seeking multiple revenue streams and in improving services, while repositories will need to make a stronger case to guarantee the flow of funding.

journals (in which TA journals arrived long before OA journals) than about the sustainability of OA journals" (point 8).

Introduction: questions and objectives

PEER (Publishing and the Ecology of European Research) has been set up as collaboration between publishers, repositories and the research community. According to the project, participating publishers made available over 53.000 manuscripts as of November 2011 that were processed by a central "PEER depot". Of these, nearly 22.500 were eligible manuscripts produced by EU authors and almost 16.700 were embargo expired and therefore available via participating repositories.

This "live lab" was created so as to make it possible to investigate the effects of large-scale deposit (so-called Green Open Access) on reader access, author visibility and more broadly the ecology of scholarly publishing. More specifically, PEER set up research projects aimed at analyzing:

- Author behaviour in self-archiving of research output and user behaviour;
- Usage of repositories and alternative access to published content;
- The cost of making research outcomes accessible via alternative means (namely journals vs. repositories) as preliminary work for assessing the economic sustainability of alternative business models in scholarly publishing and the ecology of scholarly publishing.

This report presents the final results of the "economic study"[2].

Scope and goals of this report

Following the perspective of management studies, which takes individual organizations as units for analysis, the overarching research question on which this study is based refers to the economic effect of large-scale deposit on scholarly research publication and dissemination, on the basis of the analysis of individual publishers and repositories. Addressing this issue makes it necessary first to assess cost structures associated with the two types of players; this research determines average cost of publication and distribution under different conditions, namely size, breadth of offering, and make/buy decisions, in order to determine cost drivers and their impact on the overall sustainability of different business models.

In line with existing literature on scholarly publishing, the study associates costs with specific activities, bearing in mind the relationship between key actors: authors, scholarly community, editors, publishers, libraries, readers and funding agencies. Contrary to most of the existing literature, the study analyses cost structures of individual organizations. The focus of this study is therefore to provide context for costs to specific organizations and to their choices in terms of scale and scope.

Given the multitude of ways to create value in any given industry, the term business model is often used (Chesbrough and Rosenbloom 2002) to define the combination of strategic

[2] The research team comprises Paola Dubini, Francesco di Trani, Maria Rita Micheli. The authors wish to acknowledge the help and insights of all interviewees and reviewers, the technical support of Paola Galimberti and Elena Giglia, the PEER Executive, and especially Chris Armbruster (research manager) and Julia Wallace (project manager).

decisions driving economic sustainability. In the case of scholarly research publication and dissemination, this requires that the costs associated with all the activities that need to be performed in order to provide readers with accessibility and availability of scientific research outcomes be determined and that the activities put in place to assure stability of flow, quality control of the process, accessibility and visibility for the end user, minimization of exceptions, and preservation of research outcomes be analysed[3].

Once individual cost structures have been assessed for different types of organizations (namely publishers and institutions managing repositories) and the conditions of sustainability have been identified, the assessment of the impact of Green Open Access on the ecology of scholarly publishing requires analysis of the interplay between the two models, the extent to which they compete or are complementary - and their relationship with other actors involved, namely the funders and the scholarly community. The shift from a linear information chain to an interactive communication network (Roosendaal, Geurts 1997) implies not only a change in cost structure for individual economic actors, but also a different activity distribution among actors involved in scientific research output registration, awareness, certification and preservation as well as a change in how researchers, publishers, institutions running repositories, libraries, funding agencies are interrelated (Ambruster 2009).

This shifts the nature and the distribution of externalities[4] among economic actors and the options for extracting economic value from them.[5.]

In this research we focus on the analysis of cost structures associated with research outcomes publication, document management and distribution under different business models and activities configuration, and we put these results into the broader context of research outcomes management. While this approach presents methodological difficulties associated with confidentiality and with different technicalities associated with cost calculation in different organizations, it drives the discussion on the ecology of scholarly publishing around the conditions for sustainability of different business models and on the interplay between different types of organizations.

We hope that this approach will compensate for some of the limitations of current literature. In fact, most of the available research calculates average unit costs based on aggregate data,

[3] On the difference between business models and strategy see Teece 2010.

[4] In economic literature, it is acknowledged that the exchanges of goods, services and money between parties may not only have a positive impact on the two parties involved, but may also generate outcomes for other parties as well as for society as a whole. These outcomes are defined externalities or spillovers and can be positive (economic activity generates benefits for an economic actor who does not have to bear the corresponding cost) or negative (economic activity generates costs without the corresponding benefit). The existence of externalities is a common argument used to acknowledge the imperfections in the market and to justify public intervention, in terms of norms (in scholarly publishing mandates to make publicly funded research outcomes openly available) or incentives (public funding of collaborative projects). For a review of the economic theory on externalities see Cornes Sandler 1986; on the effectiveness of different measures to control for externalities see among others Baumol 1972).

[5] Scholarly publishing and communication are collective processes. Different business models treat network externalities in different ways, i.e. the increased individual benefits deriving from the increase in the number of other users belonging to the same network. On network externalities and their impact on competition and activity configuration see Katz, Shapiro 1985.

and it is not always clear what elements of cost are included in the calculation of the overall figure. Moreover, costs are calculated at the national level, as if the business of scholarly publications were not a global one, comparing open access and SB business models as if they were mutually exclusive. The reality is that the evolution of scholarly publication and dissemination is a multinational phenomenon and that different models coexist and compete, interacting with the scholarly community and the big funding institutions. Large discrepancies exist in cost calculation, because costs are calculated in aggregate terms, and do not take into consideration the specific cost structure of specific players under specific activity configurations.

Object of analysis

Journal publishers and institutions or firms managing repositories are the objects of analysis for this study; both types of organizations select, manage, and make available different types of research outcomes. For the purposes of the research, the key research outcomes considered are journal articles. Journals are defined as exclusive peer reviewed academic publications (digital only or digital+print), gathering articles from authors in given disciplines from multiple institutions; their revenue model includes article-processing charges, subscription fees, pay-per-view, combination of subscription fees and open access article fees and other hybrid forms. Occasionally, advertising may be a source of revenue. Journals are collections of articles; as digitisation has spread, limitations in the number of articles per journal issue and in the number of journal issues during a solar year have lost importance and many journals publish articles in digital form continuously.

Journal articles are often the digital version of the corresponding paper version; this is particularly true for the research outcomes published by established commercial or not-for-profit publishing companies who started off as publishers on paper. In this case, they are established publications with codified formats and rules and treat the web mostly as a new distribution medium for a traditional text (Seringhaus Gerstein 2007). Digital only journals may have different formats, allowing for the possibilities offered by technology to perform various functions to the scientific community. In this case the web is often not a channel, but an environment in which peer reviewed research outcomes are part of a digital architecture (Seringhaus Gerstein 2006), or a digital platform linking and associating them with contents of different nature (other articles within a journal, individual articles cited by or citing the original one, comments or reviews after or before publication, advertising, databases …). The array of possible services offered to the research community within the same digital architecture includes a variety of possible formats linking articles/text to database information (Seringhaus Gerstein 2007), comments by reviewers and readers, easy links to other certified and non certified content.

While some journals, originally published on paper, have kept online the tradition of issuing new volumes at given time intervals, one of the effects of digitization is the possibility to search and access individual articles via multiple sources and to publish new articles online on a continuous basis. For the purposes of this research, this shift makes it appropriate to consider individual articles, as well as journals, as a unit of analysis, for their relevance as revenue, cost and reputation driver.

Repositories aggregate different types of non-exclusive academic outcomes. Although the focus of the analysis is journal articles, it has to be acknowledged that both types of actors may make available a wide variety of materials that include working papers, technical reports, conference proceedings, PhD dissertations, teaching materials, abstracts, research protocols, and increasingly datasets of different nature. The variety of scientific outcomes offered is relevant both for competitive purposes (as it indicates whether specific players are competing on specific subjects or products) and for cost assessment, as the variety of materials have a scale and scope impact on some cost categories. In this study, the nature of the materials archived in repositories has been taken into consideration to the extent that it includes articles. Repositories vary significantly in nature and purpose and contribute differently to the creation of such a seamless network of information, data and research outcomes. Armbruster and Romary (2009) propose four archetypes of repositories[6], on the basis of the purposes of their creation: yet, as the number of repositories and journals increases and the boundaries between different forms of research publication and dissemination blur, classification becomes more challenging and the scope of analysis widens, to include not only individual publishers and repositories, but also more complex architectures of publications[7].

Articles are the outcome of an articulated process, which is divided into logical and temporal phases: preprint (before peer review) – stage 2 (peer reviewed, accepted for publication) – stage 3 (published version). Journals manage the whole process from submission, via peer review (stage 2) to publication; the evolution of digital technology has made it economically viable for authors to make their manuscripts available to the research community early on in the publication process. Moreover, as an increasing number of funding institutions mandate that funded research be made available publicly[8], researchers have two options: either they publish their articles in journals that offer 'open access' publishing, and therefore make articles available free of charge to readers immediately on publication, or they deposit, in an institutional or subject-based repository, copies of the articles they publish in traditional journals that require a subscription from readers or their libraries[9]. The repository will then make those copies available to anyone who wants to read them, either immediately upon publication or after an embargo period designed to cushion the impact on the journal's subscription revenues. Journal articles at different stages of publication carry different value; as research is a collective process, preprints are useful within research communities to signal interest or progress in a topic, to seek feedback or cooperation, and to pre-empt competing groups of researchers. Stage 2 articles allow early circulation of quality content (in the sense that it has been already approved by peer reviewers) and increase the speed of circulation of ideas. Stage 3 articles are the definitive and certified version that can be uniquely identified, cited and preserved, although in many cases stage 2 articles are also cited. Stage 2 and stage 3

[6] Subject-based repositories are established to allow a community of scholars to share the outcomes of their work. Research repositories are put in place to record outcomes of publicly funded research projects; national repositories aim at supporting public policies concerning preservation and enhancement of national identity, while institutional repositories allow preservation and visibility of the scientific production of members of the institution and may function as internal knowledge management systems (for instance for teaching materials) or be used to check productivity.

[7] For a comparison between institutional repositories and larger scale deposits see Romary Armbruster (2010).
[8] The updated list is available in the JULIET database: http://www.sherpa.ac.uk/juliet/
[9] Should the publisher not allow deposit, some mandates require the researcher to publish in a journal allowing deposit, others will cover the cost to comply with a publisher's open choice program.

articles associate the author unequivocally with a specific journal and therefore to a specific publisher. Copyright is usually transferred to traditional commercial publishers upon acceptance at stage 2. Publishers declare their policy towards OA, and define under which circumstances the author or a third party can make stage 2 or stage 3 articles available to the scientific community outside the responsibility and control of the publisher. Unlike SB journals, authors give to OA journals non-exclusive publishing and distribution rights.

Main research questions

The specific research questions addressed in this study concern the assessment of the cost of scholarly publication and dissemination for journals and repositories. In the next paragraphs, three main business models are identified, that allow researchers to access journal articles: SB journals, OA journals, repositories of preprints, stage 2 or published articles.
In spite of the massive amount of literature available and of the growing concern about the sustainability of different business models in association with the best allocation of time and resources by interested parties, very little research is available based on actual costs to individual organizations involved with research registration, certification, awareness and preservation.

Therefore, the specific questions addressed by this study are:

- What is the overall cost incurred by publishers to make journal articles available and which activities does it encompass?

- What is the overall cost incurred by institutions managing repositories to make journal articles available and which activities does it encompass?

- What are the elements affecting the cost?

Analysing these aspects requires determining costs associated with article selection (i.e. peer review), article publication (including editing, formatting, metadata, cross referencing and uploading), costs associated with the platform hosting the research outcomes (i.e. set up costs and maintenance costs) and costs incurred to make content visible and accessible to users. All activities need to be performed to ensure certification and dissemination of knowledge (Dewatripont et al. 2006). In all three cases, make/buy decisions need to be taken into consideration, together with scale (number of documents processed, archived and managed) and scope (variety of documents archived; variety of research units involved; dispersion of research units). To assess the conditions of cost minimization, collaborative efforts in platform management or data sharing need to be taken into consideration.

- What is the cost related to compliance with PEER?

This research has been developed within the PEER project. As the project involves the collaborative effort to create a common depot, cost of compliance with PEER by individual organizations has been taken into consideration.

The three business models and their interplay are discussed in the following. A literature analysis is then presented, describing how the issues of cost of scholarly publication, preservation and distribution have been addressed, how costs have been calculated in previous research and how the issue of the management of externalities has been addressed in the discussion on the ecology of scientific publishing.

The specific empirical analysis on a sample of institutions managing repositories and publishers is then presented, discussing their cost structure and their cost drivers.

Discussion is then developed around the conditions of sustainability of specific business models; this necessarily requires taking into consideration the extent to which the three business models compete and collaborate and how they interact with the other players involved in scholarly publishing, namely authors, readers, libraries and funding agencies.

1. The interplay between publishers and institutions managing repositories

Revenue models in scholarly publishing: subscription-based versus open access

The current competitive scene shows the presence of traditional academic journals together with a growing number of Open Access (OA) digital-only journals. Both types of journals may be published by commercial publishers, university presses, learned societies or not-for-profit organizations. Publication in an OA Journal is alternative to that in a traditional academic journal and the two models are in direct competition; some publishers offer both alternatives to authors, and to institutions. For the purposes of this study, it is important to note that published articles have undergone a peer-review process and the fact they are published in a specific journal implies that the publisher certifies the existence of a review process and incurs some costs associated with the publication of the article. Other costs occur to make it possible for readers to have continuous access to the content: namely, articles need to be unequivocally identified, stored, preserved and organized so that a search by a user based on different elements (the author, the journal, the topic…) will lead her to the required content, for free or for a fee. Publishers, commercial or OA, bear these costs.

Scholarly publishing is an imperfect market. The characteristics of imperfect markets[10] include: imperfect information availability among players; differences in bargaining power among players; high barriers to entry[11]; uneven access to production technologies. There is evidence that scholarly publishing is an imperfect market: information on the price of journals or subscriptions is not readily available among players; there are no definitive ways to assess products - i.e. individual articles or journal - quality (whether measured in terms of impact factor, citations or other indicators) and there is no clear evidence of a direct relation between quality of articles and the price of journals or articles. Demand for journal subscriptions is rigid and prices inelastic, while end users may not be aware of the costs incurred by their institution nor the nature of the deal their library has signed with publishers.

Competition among publishers is high: even though the number of journals available is very high, substitution is limited by the reputation differential among journals[12]. Internal competition for visibility and impact among journals is very high and a limited number of journals enjoy a disproportionate share of attention, budget and submissions. Policies by universities providing incentives to their researchers to publish in specific lists of journals also affect competition. Competition is based on reputation; once reputation has been obtained, existing players potentially enjoy economic rewards, as the resilience of the position obtained is quite high. As reputation grows, the number and the quality of submissions, together with the rejection rate of the specific journal will increase, thus further

[10] For a recent exhaustive analysis see Calcagnini Saltari 2009.

[11] Barriers to journal publication have decreased, significantly, but barriers to access have increased: the contraction of library budgets on the one hand, growth in the number of available documents for consultation on the other make it increasingly difficult for individual journals to systematically attract a growing and loyal readership.

[12] The discussion on how such reputation is and should be measured, whether via impact factor or via other indicators, goes beyond the purpose of this research.

strengthening the journal reputation. In a world characterised by a growing quantity of research outcome, journal reputation is an effective way to filter reader attention and to pre-empt the market[13].

When imperfect markets exist, value created by the market is not evenly distributed among the actors involved. Increases in subscription prices for academic journals during the 1990s are an indicator of the existence of market imperfections to the advantage of publishers over libraries[14]. This is not to say that price increases by publishers have not been a consequence of increases in costs, or of an increase in services provided; as far as existing journals are concerned, the number of issues and number of articles published have increased – and in some cases number of pages per articles or, more appropriately, amount of information per article – (see for instance Mabe, Amin 2001; King, Alvarado Albertorio 2008). Publishers have also experienced rising costs to upgrade their technology to be able to offer journals in digital form with a higher level of service and to convert back issues of their journals. (Boyce 1998). However, the existence of market imperfections made it possible for publishers to transfer to their customers costs and inefficiencies associated with a change in processes and technologies. King and Alvarado Albertorio (2008) in their literature review on scholarly journals pricing report a series of studies showing systematic yearly price increases higher than the rate of inflation, in line with the study by Dewatripont at al. (2006:5), claiming that between 1975 and 1995 the prices of scientific journals increased between 200% and 300% beyond inflation.

At the same time, the growth in the introduction of new journals (Tenopir and King 2000), facilitated by the decrease in costs of digital publication and the slow growth in library budgets (Hawkins 1998), have contributed to an increased pressure on libraries and an average reduction in circulation of journals with a limited market potential or with lower impact.

Many markets are imperfect and the existence of market imperfections is a key justification for public intervention; by allocating resources or providing incentives, public actors intervene in the market and modify its structure. In the case of scholarly publishing, the OA movement was started by groups of researchers willing to grant dissemination of research findings, maximise visibility to publicly funded research, as well as to counteract the increase in subscription rates; it is backed by libraries and funding agencies. In this case, the argument of public intervention into an imperfect market was related to the opportunity to give equitable access to publicly financed scholarly research to scholars worldwide. From a slightly different perspective, the justification for public intervention in scholarly publishing has to do with the public nature of scientific research outcomes (David 2003). The incentives or mandates for deposit into repositories, and public set-up funding of digital platforms to not-for-profit institutions for OA journals or repositories are measures that made it possible for alternative business models to enter the industry and reduce market imperfections.

[13] McCabe 2002 suggests that the price increases experienced by journals derive from industry concentration. Authors of this study believe that market imperfections favour publishers over libraries in economic exploitation. In the presence of high competition, distribution of high-reputation journals across disciplines within the publisher portfolio may preempt competitors in libraries' purchasing budget allocation. Therefore, price increases are a consequence of market imperfections and competition, which leads to concentration, but in principle there might not be a link between industry concentration and price increases.

[14] Dewatripont et al. (2007) explain price differences among journals with market power of publishers, type of publisher, journal reputation.

Because of the relevance of reputation in determining a journal's competitive advantage, when OA journals first appeared on the market, a lively debate arose on the possible quality differential of journals in the two models (Suber 2008). However, as the number of OA journals increases and the most prestigious OA journals strengthen their reputation, thus eroding the advantage of established players, the competitive scene for scholarly publication now shows the presence of a variety of titles with different levels of reputation and robustness in the selection process both in OA and SB models[15]. On average, OA journals available on the market today tend to be digital only publications and are more frequently published by independent publishers, groups of researchers and not-for-profit organizations than SB journals (Edgar & Willinsky 2010). 10% of fully OA publishers publish two thirds of the almost 117,000 articles published; 14 publishers publish more than 1,000 articles per year and half of them use Creative Commons as licensing practice. About 8-10% of articles per year are published in fully and hybrid OA journals[16] and put OA into perspective within scholarly publishing.

In terms of business model characteristics, journals differ with regards to their revenue drivers and the exploitation opportunities expressed in copyright. Publication in traditional academic journals is typically free for the author; costs of published articles (including peer review, editorial costs, marketing and commercial costs and all costs the publisher incurs to make content accessible to the research community) are covered via subscriptions (and also via a pay-per-view model) by libraries and by individual users. In this report this business model is referred to as subscription based (SB)[17]. OA journals, on the other hand, are freely accessible to readers; article processing charges (requested by 23% of journals) are covered by institutions, funding agencies and rarely by authors.

The increasing quantity of scientific research information available, the interplay between SB and OA and the pressure on resources limiting the purchasing capacity of researchers and institutions create the context for a progressive articulation of revenue models both for SB and for OA journals. Literature and practice normally refer to hybrid models to define SB journals offering the option to their authors to provide OA rights upon publication with payment of a fee. In reality, multiple revenue models are increasingly common among both SB and OA journals. The array of opportunities to structure, relate, distribute and comment on content once it is available in digital format creates the possibility of configuring a variety of business models, each with different conditions for sustainability and leveraging on network externalities.

In 2009, 23% of OA journals asked for an article-processing fee (Shieber 2009) ranging from 400 to 2.900 USD per article[18]; as of mid-July 2011, on its website, SCOAP3 - Sponsoring Consortium for Open Access Publishing in Particle Physics – declared that most publishers in physics quote a price in the range of 1.000–2.000 EUR per published article. Most journal publishers charge article-processing charges (APC) only if the article is accepted for publication after having been peer reviewed. Article-processing charges are in principle

[15] Even reently started OA journals have good impact factors: see Giglia 2010.
[16] See http://project-soap.eu/ , and its comprehensive final report http://project-soap.eu/report-from-the-soap-symposium/. On the evolution of OA publishing see also (Laakso et al. 2011).
[17] An alternative definition, common among OA advocates, is Toll Access.
[18] See article-processing charges comparison for OA and SB journals with open choice in http://www.biomedcentral.com/info/authors/apccomparison

covered by individual researchers; increasingly, institutional membership programs are being developed by OA journals whereby organizations can enable their authors to publish in their journals without having to administer and pay "Article Processing Charges" from their individual research budget. Different models apply, ranging from prepaid membership, pay-per-use, flat rate, and waivers may apply, should the author be unable to cover article-processing charges. The Compact for Open Access Equity (COPE) is a coalition of research universities committed to establishing durable mechanisms at each institution to cover article processing for faculty publishing in OA journals, should authors have no alternative financial resources. Started in 2009 by eight institutions, it currently includes 14 relevant research institutions in the US, Canada and Europe and is endorsed by a group of Nobel laureates and prominent scholars. The explicit reference to "durable funding mechanisms" is offered as a way to reassure publishers about the duration of the commitment. The Open Access Directory (http://oad.simmons.edu/oadwiki/OA_journal_funds) lists around 45 institutions worldwide that have or are implementing some form of institutional funding of OA journals publishing fees. Universities UK and RIN (2009) describe different schemes supporting the coverage of publication of research outcomes on OA journals; OAD (http://oad.simmons.edu/oadwiki/OA_journal_business_models) and SPARC (http://www.arl.org/sparc/publisher/incomemodels) list different variations of OA journal financing, as researchers feel that article-processing charges are an important barrier for publication in OA journals.

Repositories

Journal articles and other research outputs may also be openly accessible to researchers thanks to repositories. An increasing number of authors self-archive preprints or stage 2 articles either voluntarily for visibility and reputation-building purposes, or as a consequence of mandates by their institutions or funding agencies. As a consequence of mandates, publishers too feed a selected number of repositories according to specific agreements, mostly related to embargo times. OpenDOAR (Directory of Open Access Repositories (http://www.opendoar.org) and ROAR (Registry of Open Access Repositories (http://roar.eprints.org/) show the evolution in the number of available Open Repositories.

Self-archiving in a repository (so-called Green Open Access) is complementary to publication in a journal (SB or OA). For the purposes of this research, a useful classification distinguishes between institutional and subject-based repositories.

Institutional repositories collect, preserve and disseminate the research outcomes of members affiliated to a specific institution. The affiliation might be determined by the fact that the author is hired by that institution (as in the case of faculty members of a specific university) or by the fact that she is the beneficiary of a research grant from a funding agency that has already adopted a mandatory policy, as specified in the SHERPA-JULIET database http://www.sherpa.ac.uk/juliet/ . In these cases, researchers are mandated to self-archive.

In the case of institutional repositories, the responsibilities of setting up and managing the repository are currently undertaken by public organizations or not-for-profit institutions (such as universities). Libraries, or specific sections of the institution in charge of management of

digital content are the organizational units within these organizations typically in charge of the management of the repository.

From the perspective of the institution in charge of the management of a repository, it is currently viewed mainly as a device to showcase research outputs, therefore supporting visibility and dissemination of research outcomes, while proving the amount and the quality of the research performed by affiliated researchers. As an increasing number of institutions is mandating its affiliated researchers to feed their own repositories or third-parties repositories, repositories may become comprehensive sources of information for the scholarly community. Potentially they will also become an asset for reputation building as well as for internal purposes (such as knowledge management and support to increase research effectiveness and efficiency of research activity, assessment of research productivity). The possibility of using the repository as a strategic asset depends on how it is designed and structured and the extent to which it is used. Depending on the amount of relevant research output, ease of use and accessibility by users, repositories may become research-dissemination channels partially overlapping with publishers' user interfaces.

The author may upload her research outcomes on a repository for three main reasons: early visibility, willingness to engage in a debate with peers, and mandate from her research institution or a funding agency. From the author's point of view, presence in a repository may increase their visibility (Swan 2010), particularly to materials (such as PhD dissertations) that would otherwise be very hard to find. For funding agencies (as well as academic and research institutions adopting mandatory policies), presence in an institutional repository allows open access to publicly funded research and a higher return on investment (Houghton 2010). The diffusion of repositories therefore may increase the possibilities for individual research outcomes to be visible, accessed and discussed, while at the same time offering access opportunities to researchers in institutions with limited funds to buy access to SB journals.

Institutional repositories are typically financed with public funding. They best serve the community if and when they contribute to easy, fast and free access to research outcomes. This may occur in different ways:

- Because the repository is recognized as a comprehensive source of information due to its size, accessibility, reputation. It is therefore an entry point for researchers or a source analysed on a regular basis, also because it provides free access to full text articles.

- Because the repository is well connected to other sources of information, so that a search will determine visibility for documents hosted in the repository.

- Because the repository hosts unique content that is otherwise unavailable or difficult to access.

Thanks to the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) used by almost all repositories, harvesters have been implemented that gather the metadata and build a collective database of bibliographic references linked to full text when available. Good examples are Base, CiteSeer, OAIster and Scirus.

Researchers may also opt for preprint and post-print archiving into subject-based repositories

(such as SSRN or ArXiv); this is another complementary form of journal publication, allowing early visibility. Not surprisingly, subject-based repositories have emerged in domains with a long-standing tradition for exchanging pre-prints and where speed of publication is particularly relevant, in physics and computer science, for instance.

A very interesting example is ArXiv, an electronic archive of freely accessible research preprints (Ginsparg 1997). Started in 1991, it has since become an indispensable tool for researchers mostly in physics and mathematics. To participate, authors need an affiliation with a recognized academic institution or an endorsement from an established author. Given the relevance within the community, the transparent process of rating articles, the option to cite the contribution before journal publication (Gentil Beccot et al. 2009) and relatively simple organizational solutions related to certification have allowed the archive to become de facto an alternative to publication (Boldt 2010) and therefore a competitor to SB and to OA journals[19]. ArXiv is a quite unique case of a repository with such a reputation and market share to be close to a publishing alternative.

While institutional repositories are financed with public funding, subject-based repositories may be developed by private companies, by teams of researchers within a research institution, or be part of a broader platform within an institutional repository. Development of interoperability standards and collaborative projects have made it possible for repositories to grow their content base and increase their potential to become a stable player in the scholarly research dissemination industry.

The interplay between journals and repositories

As Willinsky (2009) notes, the scholarly publishing and dissemination field is stratified and different interacting models exist. From the point of view of visibility and content accessibility to research outcomes, the relationship between repositories and publishers of OA journals is very friendly, as they mutually support each other. An effective repository – i.e. a repository that is recognised as a comprehensive source of information and based on a platform allowing easy visibility or retrieval of information from several sources – supports OA journals in their quest for visibility. In turn, as OA journals are freely accessible by definition, they increase the attractiveness of repositories, which are particularly useful when they provide access to full text. As Tenopir et al. (2009) point out, indexing, citation counts and online access to individual articles are increasingly critical elements driving searches, as opposed to the traditional way of reading selected journals.

In this respect, the relationship between repositories and publishers of SB journals is more problematic. Publishers sell content, whereas repositories make it available for free. In principle, an ideal setting for the publisher would be to view the repository as a pure marketing tool, signalling the existence of the article within the journal, but then referring to the publisher's website for purchasing. On the other hand, part of the value associated with the repository is often correlated with the free availability of full-text certified content. Institutional repositories can increase the number of articles available through mandatory self-archiving of author copies or mandatory deposit by authors or publishers of research

[19] Yet, according to Suber (2009) no SB journal in physics has closed since 1991, despite the presence of ArXiv.

outcomes financed by an increasing number of funding agencies. Additionally, negotiation with publishers on embargo-time reduction and harvesting can also assist. As far as subject repositories are concerned, the possibility that authors self-archive their research output depends also on the quantity and the quality of the content published, on the number of readers, and on the variety of services offered by the repository to the community.

Björk et al. (2010) report that the overall share of OA references (either from pure Gold OA journals, delayed OA journals or openly accessible articles from repositories of a different nature) is approximately 20,4% of total articles published in 2008. While most of the open articles come from pure OA journals, Way (2010) found that 27% of articles in a sample of 922 articles published in 2007 in 20 top journals were available as OA journals, mainly in subject-based repositories and on authors' personal websites. Björk et al. report that 38% of articles they analysed were exact copies of those published, 46% stage 2 and 15% preprints.

Although alternative publishing models have spread with different intensity across disciplines (Kling, McKim 2000), and not as fast as proponents were predicting (Björk Hedlund 2009; Basefsky 2009), they are currently a stable part of the competitive landscape in scholarly publishing. The Directory of OA Repositories (http://opendoar.org) lists more than 1.780 repositories, while the Directory of OA Journals lists over 5.950 titles (http://www.doaj.org/), which are part of the 25.000 peer-reviewed journals and refereed conference proceedings (Gargouri et al. et al. 2010)[20].

All these players face significant costs to archive, make available, preserve and make visible the content they manage to scholarly communities worldwide. The costs associated with each model and the conditions for sustainability are the object of investigation in the following paragraphs.

[20] Currently, DOAR lists 2.144 repositories and DOAJ 7.299 journals.

2. Literature review analysis

Relevant literature for the purpose of this research deals with three main areas of analysis[21]:

- Value creation in scholarly research and sources of competitive advantage for the players involved. This necessitates discussion of the key activities associated with scholarly research publication, distribution and communication.
- Cost assessment associated with different activities, to different actors and to different business models; we were particularly interested in contributions assessing costs associated with the certification and publication of content (i.e. peer review, and publishing), costs associated with the platform hosting the research outcomes (i.e. set up and maintenance costs) and costs associated with the archiving of documents (metadata, cross referencing, uploading).
- The economic sustainability of different business models within the dynamics of scholarly research.

Scholarly publishing has witnessed significant structural changes in the past twenty years; the research team decided not to take into consideration any specific contribution dealing with scholarly publishing economics prior to Getz 1992. Most recent contributions dealing with cost assessment and the interplay between different actors involved in scholarly publication dissemination and communication have been identified starting from the contribution of King (2007)[22].

Value creation in scholarly research

There is a rich body of literature addressing the issues of value creation of scholarly publishing; within content industries. Indeed scholarly publishing has been one of the first segments to experience the challenges of digitization, and publishers have faced the problem of managing production and distribution of paper as well as digital versions of their journals (Marks Duranceau 1995; Odlyzko 1997, 1998; Boyce 1997; Fisher 1997; King Tenopir 1998; Bot et al. et al. 1998). The impact of digitization on production costs (Getz 1992) is the focus of analysis of several studies; authors stress the different impact of scale and scope economies on production of digital journals (Duranceau 1995) and explore the possibilities for disintermediation and outsourcing. Another series of relevant articles for the purposes of this research deal with the identification of key activities associated with content publication and dissemination. For instance, King (2004) discusses the relevance of article-processing costs, while Getz (1992) focuses on storage and typesetting activities. Fox (2002) identifies

[21] In order to classify systematically the most relevant articles on these topics, we started from the analysis of the annual edition 2008 of the bibliography by Charles W. Bailey Jr. This work offers selected English-language articles, books, and other printed and electronic sources aimed at understanding the process of scholarly electronic publishing on the Web. We focused on the part devoted to Economic issues and we analysed each article presented to decide which to include in our literature review. The section contains 98 articles and we kept 48 for our purposes. For the oldest articles, we checked whether a more recent version was available and we made comparisons among the different versions.

[22] This work has a totally different approach from the one followed in this paper, but it was considered a comprehensive work with which to begin. A more comparable approach is followed by Houghton et al. (2009).

activities such as intake, processing and preservation (maintenance hardware and software, updates, inspection) that become increasingly important as publishers shift to digital and have fixed costs. Rowland (2002) compares peer review costs in traditional as well as electronic publishing. The report of the Wellcome Trust (2004), through interviews with publishers, identifies several activities associated with the journal-publishing process and their impact on publishers' total costs. Activities can be summarized as follows: refereeing (22% of the total cost), editorial and typesetting (33%), subscription management (7%), physical production and distribution (23%), sales and marketing (13%), and promotion (2%). Willinsky (2005) explores the savings associated with automated digital journals, compared to traditional print journals. He identifies different stages, multiple agents involved, and different processes for automated and assisted journal management.

Digitization affects not only the way journals are produced and distributed and the corresponding cost structure, but also offers alternative configurations. As content is separated from the medium of paper, the commercial relationship between publishers and their traditional customers, i.e. libraries, changes. Subscription to physical products is gradually transforming into a service and libraries may be charged differently depending on the bundle of products to which they subscribe. Digitization allows for price discrimination (Varian 1996a and b) and therefore a maximisation of revenue potential, as different markets can be served at different prices. For instance, different bundles of journals may be sold at different prices to individual libraries or to consortia; participating libraries may be charged different prices depending on the use; moreover, the same units of outputs may be sold to different classes of customers at different prices. Moreover, different prices may be set for access and for reproduction, or for access to current and past issues of specific journals. Halliday and Oppenheim (1999) explore several economic models for electronic publishing, focusing on distribution activities and highlighting the different costs associated with them. They explore the price evolution, correlating it with subscription and technology evolution and considering both fixed and usage costs (e.g. re-shelving). Also the time spent in performing the different activities for different actors has been included in the costs calculation. The availability of content in digital form allows for a much for effective content bundling and price discrimination[23] by publishers (Peters 1999; Chen et al. et al. 2001; Montgomery 2002). One of the effects of digitization and of technological evolution is the multiplication of revenue models and associated cost structures. For instance, Hedlund, Gustafsson, and Björk (2004) develop a model to map the activities of all involved stakeholders and to clarify them. The model is then used to estimate empirically the costs of alternative business models.

The evolution of the subscription model, the option to charge for individual downloads, the development of "big deals"[24] agreements replacing the "pure" subscription model are explored, by taking into consideration publishers and libraries' point of view (Frazier 2001; Gotten and Sanville 2004 Schonfeld et al. 2004; Lewis 2004; Hahn 2006; Dewatripont et al. et al. 2006). Depending on how deals are structured, the impact of price discrimination may

[23] Digitization significantly impacts the economics of the so-called information goods. For a comprehensive analysis see Shapiro Varian (1999).

[24] Under the "Big Deal" agreement, referring to print and electronic versions of journals, contract. a library or a library consortium enters into a long-term arrangement to get access to a large electronic library of journals at a substantial discount in exchange for a promise not to cut print subscriptions (the prices of which will increase over time). Edlin Rubinfeld 2004.

be positive for the publisher, who can maximise profitability for its title portfolio, it may alter the relative power of publishers of different size and for journals with differing reputation (Jeon Menicucci 2006); impact on libraries is mixed, as – depending on the nature of the deal – they might experience reduced degrees of freedom in choosing the journals to subscribe to (Frazier 2001), but also may enjoy a wider choice of titles for the same price (Gerhard 2005). Library consortia signing deals may enjoy advantages from cooperation, particularly if searched material is similar and if the amount of journals accessed too is similar across participating libraries (Dewatripont et al. et al. 2006).

Another set of studies addresses the impact of digitization on libraries cost structure. Several effects of digitization occur on libraries processes, cost structure and sources of competitive advantage. Libraries offering configuration changes, as researchers increasingly access content remotely and new challenges emerge, as the need for long term preservation of digital copies (see for instance Bowen 1996; Fox 2002). Montgomery et al. (2002) focus on the costs for libraries, comparing print and digital journals. Costs related to the development of the collection are identified and activities related to communication, physical handling, record creation, maintenance and references are explored.

In these papers, traditional scholarly research publication is mostly described as a linear process. Once a manuscript is produced, it is submitted to a journal for peer evaluation and possibly for publication. In parallel, it may be posted by the author in a subject based repository. The publisher (or whoever plays the publisher role) then organizes peer review. Authors may be offered a variety of services such as tracking of the manuscript, visibility on the comments by reviewers, and so on. Once the review process is completed and the manuscript is considered of adequate quality to be published by a journal, a series of activities need to be performed in order to publish it (formatting, typesetting, and copyediting) and to make it easily accessible and searchable (metadata creation, abstract and keywords). If the journal has a printed version, it is physically duplicated and distributed to libraries; alternatively, it is made accessible via web through an interface that may provide a variety of services to the user beyond mere accessibility: link to related content or to references, comments by users, information on usage and citations. The flow of certified content from authors to users is organized and managed by publishers; libraries are the prime customers of publishers. Digitization affects how processes are performed, offering reconfiguration and impacts on journal publishing cost structure, changing the impact of scale and scope economies and the incidence of different cost categories. All these activities generate costs and contribute to value creation, and the overall value created has to do with content certification on the one hand (the most valuable content is the one that is published in the most reputed journal, which is generally the journal where selection is harsher) and control of distribution on the other. Revenue maximization is related to unit prices and market size; profitability obviously has to do with process efficiency (thus the number of studies analysing efficiency gains in the transition to digital journals), but market inefficiencies favour established players over new ones, publishers strong in disciplines characterised by big researcher communities, and higher impact journals.

Cost assessment

A second relevant stream of literature deals with the costs associated with different business models and with the quality of research outcomes. In this framework, the emergence of OA journals backed by public funding can be viewed both as a way to reduce market imperfections and a way to reduce the pressure on library budgets. By allowing newcomers to enter an industry strongly controlled by incumbents, the possibility for incumbents to automatically transfer cost variations to prices is limited and competition increases, assuming that newcomers can prove they can offer journals able to quickly achieve a comparable reputation. Not surprisingly, there is a stream of research concerned with quality of OA journals on the one hand (see for instance McCabe and Snyder 2005, Jeon and Rochet 2007) and with citation impact on the other (Lawrence 2001). The mainstream opinion in the information science literature is that OA increases the number of citations received by scientific papers and that this effect is quantitatively important (see for instance Eysenbach 2006 for the completeness of the study from a modelling point of view). These findings have been challenged by other authors (see for instance Davis et al. et al. (2008)) who suggest that the effect of OA is heterogeneous across journals.

Clarke (2007) aims to establish a cost model for alternative types of digital journals. A detailed scheme is proposed that underlines which agents are at an advantage and which at a disadvantage under different schemes. Differences among for-profit and not-for-profit publishers are explored. Recent contributions addressing the issue of cost of scholarly publishing claim that OA significantly reduces costs of publication. Table 1 synthesises costs per published article as calculated by Houghton et al. (2009) for toll access publishers, in pounds and in USD at an exchange rate of 1,33. Comparable data for OA journals lead authors to claim that the cost per article of a digital-only OA is 1.524 GBP (2.026 USD).

Table 1. Cost structure of commercial publishers (Houghton et al. 2009)

	GBP	USD
Peer review	344	458
Internal		
Peer review external	1.390	1.849
Publication costs		
Editing, composition, typesetting	152	202
digital only		
Administrative	30	40
Hosting	200	266
Marketing and helpdesk	170	226
Subscriber driven costs	51	68
Total	2.337	3.109

One merit of this study is a detailed definition of activities involved in the process of content selection and certification, publication and distribution. However, five elements call for caution:

- Data are obtained by referring to previous studies (Tenopir & King 2000; Waltham 2006) and updating value at 2007 prices, not on interviews with players;
- It is not possible from Houghton (2009a) to infer the impact of publisher scale and of make/buy decisions on cost levels. Therefore, it is not possible to determine whether the cost per article is a gross average, or whether it refers to a publisher like Wiley (with a declared archive of 4 million articles in 2010 and 120.000 new articles per year added to its platform) or BMJ, with a bit more of 824.000 articles and a yearly addition of nearly 52.000. Moreover, make/buy decisions impact on the incidence of fixed costs on the overall cost. Houghton calculates a full cost per article, which is heavily influenced by the overall size of the publisher and by the incidence of fixed costs[25]. The bigger the size of the publisher and the higher the incidence of fixed costs, the lower the unit cost.

[25] Full costing is a cost calculation technique whereby the average unit cost of a product is obtained by dividing total costs by the number of units produced; in this case, cost is calculated by dividing total costs by the number of articles published. The total number of articles and the incidence of fixed costs both have a significant impact on average full cost. Given a similar incidence of fixed costs over variable costs, full cost tends to be lower for bigger publishers. Make/buy decisions affect the incidence of fixed costs on total costs.

- Similarly, average full cost for commercial publishers includes costs for printed versions of journal, whereby full cost for OA journal is calculated for e-only journals.

- The difference in cost for OA journals and SB journals is mostly attributed to a lower cost of platform, associated with a lesser need of protection. However, it is not possible to determine from the study the cost savings associated with the fact that OA does not require content protection. Also it is not possible to determine the wealth of service and revenues opportunities deriving from the fact that content is openly accessible.[26]

- The full cost of commercial publishing normally includes the cost of the proprietary platform developed to host the articles; in contrast, many OA journals are published on open source software based platforms developed and managed by third parties, often research institutions, and pay modest fees to access platforms (Edgar and Willinsky 2010) developed and maintained with public funding. Determining the cost of these platforms goes beyond the purpose of this study: King and Alvarado Albertorio (2008) report an initial grant of 9 million USD for PLoS.

Thanks to platforms designed to host OA journals or to minimise the costs of operating and running a repository, barriers to scholarly publishing and content dissemination have decreased for groups of researchers interested in setting up a journal. As Fisher (2008) points out, publishing a journal can cost between 4.000 and 5.000 USD a year per journal. Edgar and Willinsky 2010 show a wider range of cost possibilities, but acknowledge that the cost of managing a journal may be very limited.

As far as repositories are concerned, Swan (2008) reports for DSpace a set up cost of the order of 2.4-2.5 million USD, of which 1.8 million USD is specifically for technical costs, 3 FTE (full-time equivalent) staff (200.000 – 300.000 USD) and 400.000 USD for system equipment. Swan also claims yearly operating costs to run DSpace of 285.000 USD (225.000 USD for staff, 35.000 USD for system equipment and 25.000 USD for general expenses), which is the same amount estimated by Barton et al. (2003).
Swan (2008:37) infers that an institution can set up a repository with less than 10.000 EUR for platform development, to either get professional service to integrate a software like Eprints, or to pay for internal resources to adapt DSpace; management costs of feeding and maintaining an institutional repository may be of the order of very few FTE staff, provided organizational (such as mandatory policies (Darnton 2009)) or technological solutions (such as harvesting from subject-based repositories (Proudman 2008;56)) are put in place. Part of the costs associated with repository management are associated with the implementation of mandates within the organization, quality control on the documents uploaded, marketing of the repository within and outside the organization and upload of documents instead of authors. Numbers are approximate and incomplete, as the value of Swan (2008) is an average

[26] The case of BioMed Central is particularly interesting in this respect. The publisher's website is structured as a portal and has advertising as a revenue stream. Should an OA publisher (or a collaborative project based on OA content, as is the case for SciELO - Scientific Electronic Library Online, a cooperative project of electronic publishing aimed at reducing the impact of distribution and dissemination barriers for scholarly content produced in Latin America) have a sufficient mass of articles and reputation to be able to attract readers, then OA allows them to open up advertising as a revenue stream. Obviously, the size of the research community matters: not surprisingly, Medicine and Life Sciences publishing in general shows a much higher degree of revenue gained through advertising than other disciplines. This holds true for SB journals just as for OA.

of data gathered at different times, in different currencies and is based on estimates about staff and overheads that need to take into consideration massive sunk costs. Yet, these contributions do make the point that journals and repositories may be set up with limited upfront investment, allowing even small communities of researchers to share the outcomes of their research at negligible cost. These studies also imply the presence of organizational interventions (the costs of which are in part very hard to determine) to grant a smooth and constant upload of articles and quality control.

Houghton (2009a), Fisher (2008), and Swan (2008) suggest that differences in publication costs can be attributed to the higher efficiency of OA models. The existence of journals like PLoS prove that peer-reviewed OA journals can be sustainable. However, little systematic empirical evidence exists on cost structures of different players involved in scholarly content publication, dissemination and preservation to determine the best resource and activity allocation of public funding to simultaneously maximise scholarly publishing quality, visibility and preservation, notwithstanding the presence in the market of both SB and OA journals.

Edgar and Willinsky (2010) offer probably the first systematic study on the cost of OA publishing. Their survey of 998 scholarly journals published using the Open Journal Systems (OJS) platform[27] shows the effectiveness of the model to support publication of certified scholarly research outcomes in times of budgetary constraints. The study takes journals and not articles as the unit of measure and does not provide data on the number of articles per journal. As Table 2 shows, average operating cost for an OA publisher can be kept at relatively low levels. Note though the differences in the results compared with Fisher (2008), suggesting that the two studies do not take into consideration the same array of costs. Summing up the items listed in Table 2, it appears that an OA digital-only publisher can operate with a yearly budget of about 50.000 USD, (about 60.000 USD if the journal is published both in print and digital versions) without overheads (rent, equipment, energy, general expenses).

[27] "OJS is one of a number of open source journal management systems (see Cyzyk & Choudhury, 2008, for a review of comparable systems). It is being used by approximately 5.000 journals, has had 19 upgrades since it was first made available in 2002, and is now available in 20 languages." (Edgar and Willinsky 2010:2)

Table 2. Annual journal expenses (USD) by number of journals

Expense category	% of journal with cost=0	Mean/USD
Editorial	79%	6.442
Management	73%	9.053
Promotion	88%	1.991
Publishing*	76%	8.342
Technical**	81%	4.746
Platform subscription	96%	12.728
Print edition***	61%	11.989
Other	75%	5.447

* Copyediting + article layout + proofreading

** Website + technical + customization

*** Journal printing + postage

Source: adapted from Edgar and Willinsky (2010), Table 14, p. 14.

As Table 2 shows, the percentage of journals with no expenses associated with the main cost items is very high; as Edgar and Willinsky (2010) suggest, this model is particularly relevant for self-publishers, i.e. groups of researchers taking advantage of the opportunities offered by digital technologies to disintermediate commercial publishers. Evidence for this is the high percentage of unpaid work associated with journals publication in the sample considered in the study (Table 3) and the heavy involvement of authors and editors in the actual journal publication (Table 4).

Table 3 Basis of staff participation in journal roles by number of journals

Role	Number	% no fee paid
Editors	829	77%
Journal managers	784	63%
Peer reviewers	807	90%
Graphic designers	751	35%
Technical support	765	44%
Clerical support	713	36%
Promotional work	708	44%

Source: adapted from Edgar and Willinsky (2010), Table 9, p. 11.

Table 4. Involvement of different organizational roles in the publication process by number of journals

Staff position	Copyediting	%	Layout	%	Proofreading	%
Journal editor	695	76	532	58	646	70
Staff	299	33	418	46	309	33
Volunteer	168	18	157	17	176	19
Author	274	30	139	15	454	49

Source: adapted from Edgar and Willinsky (2010), Table 8, p. 10.

In spite of this, the economic sustainability of publishers following this model is not yet systematically guaranteed. 50% of journals analysed break even, 21% are profitable without accounting for overheads and financial costs, 28% report a loss. This may be partly due to the average low age of journals considered in the study.

Charging article-processing fees is a way for OA journals to cover operating costs. Dallmeier-Tiessen et al. (2011) study of 22.977 authors who published at least one article in an OA journal reveal that 50% of the respondents had published with no fee, while in the other cases, the median value ranged between 500 and 1.000 EUR and was paid by the author in 12% of cases. The highest percentage of articles published with no fee are in the fields of humanities and the social sciences.

Economic sustainability and the ecology of scholarly publishing

Given the specificity of scholarly research publication and dissemination, the transformation it has been facing in the past twenty years (from print publishing to digital publishing to OA journals to repositories), the variety of possible activity configuration, the interplay between different actors (with funding agencies, libraries and group of researchers taking strategic initiative to reshape the industry and some commercial publishers reacting quite actively in articulating their offering), no single study on the comparison between business models seems to satisfactorily prove superiority of one model over the other. By supporting an alternative model to SB, funding agencies and institutions have made it possible to reduce market asymmetries and allow entry into the market for new OA journals. As Willinsky (2009) notes, digitisation has made it possible for groups of researchers to develop sustainable low-cost ways to allow scientific communities to publish and be visible without having to pay high article-processing fees, but at the same time without necessarily giving up on reputation. At the same time, while there is a growing body of literature demonstrating the viability and relevance of collaborative and open projects, sustainability still has to be proven in a systematic way. Some OA journals have managed to grow and establish themselves as solid competitors to SB journals in visibility and reputation; some SB journals have revised their revenue models and started offering OA options; in both cases, shrinking budgets tend to favour high reputation journals. Low fee or free OA publication platforms allow self publication by groups of researchers or learned societies (see Cooney McQuat Busch and

Kahn 2010). Boundaries among players change, as some activities originally performed by publishers are transferred to other players (Bergstrom 2001; Carr, Harnad 2005; Ambruster 2009). The political pressure of the OA movement (BOAI 2002; David 2003) significantly contributes to the opening of the competitive space to include OA journals and repositories and to the fuelling of resources specifically aimed at building digital infrastructures to host preprints and post-prints (Barton Walker 2002). Scholarly publishing is no longer simply the transformation from print to digital, rather it is a new competitive space incorporating a variety of business practices (Benkley 2002).

All these elements led a variety of studies to examine scholarly publishing within the broader process of scholarly research production and dissemination. Therefore examining the interplay among players, the exploitation of externalities by different actors, the non-economic value created by the faster and more effective flow of ideas among researchers around collaborative platforms, will ultimately reveal the contribution to society of alternative models of research certification, publication and communication.

In this respect, some studies focus on the identification of the key functions of scholarly research outcomes, namely registration (the author is recognised as the person who carried out the research and made specific discoveries), certification (the peer community acknowledges the author's research and the appropriateness of the findings), awareness (the research outcomes are communicated and made available to the scientific community) and archiving (so that the research is retained for posterity) (Roosendaal and Geurts 1997).

Depending on how these activities are distributed among the actors, different costs and benefits occur, not just for the individual players involved, but also for society as a whole (Houghton 2001). Fisher (2008) identifies a group of seven activities which are related to scholarly publishing and that generate different kinds of costs: acquisition and management of intellectual goods, peer review activities, editorial management, creation of a working platform, formatting and publishing, displaying and storing the final product and for indexing, archiving and sharing. Morris (2005) analyses the costs of scholarly publishing considering the whole context of the total cost of research communication and focuses on costs borne by libraries and publishers too, such as preservation. By acknowledging the possibility that OA might not be the solution to better awareness and to more efficient research communication, she suggests that conditions for sustainability and continuity be analysed. Joint (2006) analyses the costs related to Green Open Access and the role of libraries. A particular focus is on the activities of metadata creation and digital preservation. According to Björk (2007), scientific communication is an interconnected information system and the activities of different actors are mapped in connection with the interlinked steps of the process. Houghton et al. (2009) develop a comprehensive economic model of scholarly publishing under different business models. The activities which constitute the scholarly publishing processes are: reading, writing, peer review, editorial activities, editorial board activities, preparing grant application, reviewing grant applications and publisher costs. For the different phases costs and benefits associated with different business models are identified and quantified.

Green OA allows for registration, awareness and archiving - performed by researchers and research institutions and disintermediates conventional publishers (Prosser 2005); the advantages of self-archiving and Green OA are systemic and not just confined to individual players (Suber 2005). Deposit in OA archives increases the frequency of citations of articles

(Harnad and Brody 2004; Gargouri et al. 2010) provided they are well indexed and linked to search engines such as Google Scholar.

As a consequence of digitization, the four key functions of scholarly publishing can be performed by different players (Dewatripont et al. 2006; Ambruster 2009). One of the functions of repositories is to separate certification from dissemination: journals certificate quality, whereas repositories disseminate and create awareness to market segments that would be constrained by high subscription prices. Edgars and Willinsky (2010) show that by publishing via publishing platforms like OJS, groups of authors can gain an economic advantage in self-organising peer review, bypassing SB journals and using repositories as dissemination channels. The same can occur in learned societies journals, particularly the smaller ones[28], who suffer from the cancellation of library subscriptions and might therefore consider disseminating via repositories instead of SB journals (Cooney Mc Quat et al. 2010). Costs of term preservation may be covered jointly by publishers and libraries participating in collaborative projects such as CLOCKSS[29].

In order to assess the impact on society of the interplay of different models leveraging the four key functions of scholarly publishing, some studies develop scenarios at the macro level on the possible costs and benefits for society associated with the growth of OA. For instance, Björk and Hedlund (2009) envisage two possible ways for scholarly publishers to shift to OA, based on the underlying assumption that OA is the scholarly publishing model that maximises awareness on the one hand (Harnad et al. 2004; Gargouri et al. 2010) and maximises efficiency on the other (Houghton et al. 2009). The two authors explore the possibility of either a drastic move from SB to OA (Suber 2007) or a gradual one, allowing for the development of alternative pricing schemes for authors and institutions.

Differences in reputation among journals (and therefore the associated price elasticity), in size of publishers, and in publishing and research dissemination patterns across disciplines contribute to make scenarios difficult to envisage. Some studies quantify savings at the national level associated with a systematic promotion of OA across institutions: studies of this type typically present different scenarios, encompass all players involved and calculate overall costs based on secondary data. The modelling effort is enormous, due to the number of variables that need to be taken into consideration, but necessarily the underlying hypotheses are very strong and imply a significant level of coordination among a large number of players and a quick reduction of information asymmetries among research communities, which in practice are still very large. Among these studies see for instance Houghton Steele Sheehan (2006), Houghton (2009a).

Other policy-oriented studies (such as RIN 2008[30]) are mostly concerned with the collective savings associated with the shift to digital-only publication and with the distribution of costs and savings among activities and players in the event that one publishing model (i.e. OA) predominates.

Besides the quality and sophistication of different models, these studies are important as they

[28] Crow (2006) shows that over 97% of learned society publishers publish fewer than three journals, with 90% publishing just one.
[29] http://www.clockss.org/clockss/Home
[30] http://www.rin.ac.uk/our-work/communicating-and-disseminating-research/activities-costs-and-funding-flows-scholarly-commu "Activities, costs and funding flows report".

highlight the political dimension of the evolution of scholarly publishing. Changes in the industry have been determined not only by user behaviour and technological change, but also by the introduction in the system of policies (as in the case of mandates by big funding agencies) and incentive systems (as in the case of research institutions mandating their researchers to self-archive in institutional repositories). What these studies tend to overlook is the process dimension and the network effects of these policies. As Dewatripont et al. (2006:44) highlight, the interdependence of players is very high, as well as path dependent, thus making any shift less deterministic than some of the scenario-based studies might suggest.

The last important contribution of studies concerned with the overall impact of OA on scholarly research publishing and dissemination refers to the issue of sustainability of different business models. OA was originally supported to balance market imperfections and information asymmetries in subscription prices. As a consequence, Gold OA journals emerged as publishing alternatives to SB journals, repositories offered readers the possibility of accessing research outcomes without pay-per-view (directly or indirectly) and provided OA journals with extra dissemination channels. As the scientific community is becoming more aware of the differences between the various models in terms of accessibility and copyright management, repositories more established within specific communities may offer publishing platforms, thus reuniting certification and awareness functions.

Once OA journals and repositories established themselves as viable alternatives to SB journals, the three business models increasingly competed on the four functions of the increasing amount of research, the strain on library and funding organizations budgets, the pressure on authors to publish or perish, and the finite amount of time devoted to reading and researching.

Even though price-aware scientists appreciate OA journals as readers, OA is quite low on authors' lists of preferences in deciding where to publish, as the promise of a higher visibility and citation does not systematically hold true and research budgets may be limited.

However, as SOAP 2011 suggests, the tipping point in some disciplines may be relatively close, as an increasing number of scholars recognise the opportunities afforded by OA. Funding and the perceived lack of high-quality journals in many disciplines are the main barriers to fill the gap in the behaviour of researchers as readers and as authors.

It has to be noted that the increase in article-processing fees in OA journals has taken place at a much slower pace than in SB journals. Even when impact is high and submission flow is satisfactory, the sunk cost of running an OA journal by a group of researchers may threaten its continuity, given that the journal is managed on a voluntary basis and time devoted to research publication is not devoted to research (Cavaleri et al. 2009). In a highly price inelastic industry, the evidence of a price differential associated to OA may induce researchers and institutions to favour OA journals over SB, for the same level of reputation.

In order to encourage publication in OA journals, the resistance of authors to cover publication costs may be overcome by negotiating with research institutions or funding agencies a dedicated budget to pay author fees for all articles submitted to OA only journals (Shieber 2009). Indeed, this is already happening (Dallmeier-Tiessen et al. 2011), although in

this case, research institutions need to finance OA publication. The issue of the sustainability of OA journals therefore becomes important, as research institutions or groups of researchers may find themselves involved not only with research production and dissemination but also with journal publication.

A similar problem of role assignment arises for funding agencies that not only fund research but also the publication process itself (Imboden 2009). Within the scholarly research field, they are likely to be most concerned with the fact that OA is not currently replacing SB publishing models, but rather in part duplicating it. At the same time, the emergence of OA calls more careful attention to be paid to resource allocation between research and dissemination and between the different publication and dissemination options (Kaufman Wills 2005).

As the key functions of scholarly publications may be distributed among players in a competitive environment that is increasingly articulated, it seems appropriate to turn to individual players and understand the conditions for their sustainability, as a prerequisite to inferring the future scholarly publishing scene.

3. Method and data

The sample in this study consists of 22 organizations involved with journal article publication and dissemination[31]. The organizations are listed in Table 5 and have been analysed via direct interviews and looking at information provided in annual reports, public presentations, descriptions in articles and documents cited in bibliography[32].

Table 5 The sample

Publishers	**Institutions/firms running repositories**	**Open access publishers**
BMJ Group	Cornell University *	BioMed Central
Cambridge (CUP)	*CSIC*	Hindawi
EDP Sciences	**GESIS**	*PLoS*
Elsevier	**Göttingen University**	
IOP Publishing	**Inria****	
Nature (NPG)	**Max Planck Society ****	
Sage Publications	NIH ***	
Springer	SSRN	
Taylor & Francis	*Università degli Studi Milano*	
Wiley-Blackwell		

* with respect to ArXiv
** with reference to central institutional repository eDoc
*** with reference to PubMedCentral US
**** with reference to HAL and is run jointly by CNRS and Inria

Publishers include commercial companies and not-for-profit organizations publishing academic journals. They might have a catalogue of OA journals, but their revenues from journals rely heavily on subscriptions paid by libraries and institutions and their competitive advantage traditionally derives from reputation and/or the number of journals in their catalogue. They were typically born as print publishers and have gradually increased their offering to include digital versions of the journals. In some cases, the shift to digital journals has made it possible to increase the number of articles published and to shorten the time to publication. Journals varied significantly in terms of reputation, associated with their rejection rates. Publishers in the sample also differ significantly in terms of the number of journals published and the scope of domains covered. Table 1 in Annex 1 on page 67 describes the companies in the sample.

Three active OA publishers showing high growth rate in the number of articles offered and a remarkable size were included in the sample: BioMed Central, PLoS and Hindawi. These publishers were taken into consideration, because of their size, growth rate and reputation, their competitive advantage derives both from reputation (attracting authors), number of

[31] In bold are institutions that are part of the PEER project, institutions in italic were analysed based on published documents, but representatives from the organization were interviewed once, while information on the remaining institutions derives exclusively from publicly available material.
[32] Data are updated to the end of 2010

articles published, accessibility and availability, which are elements attracting readers.

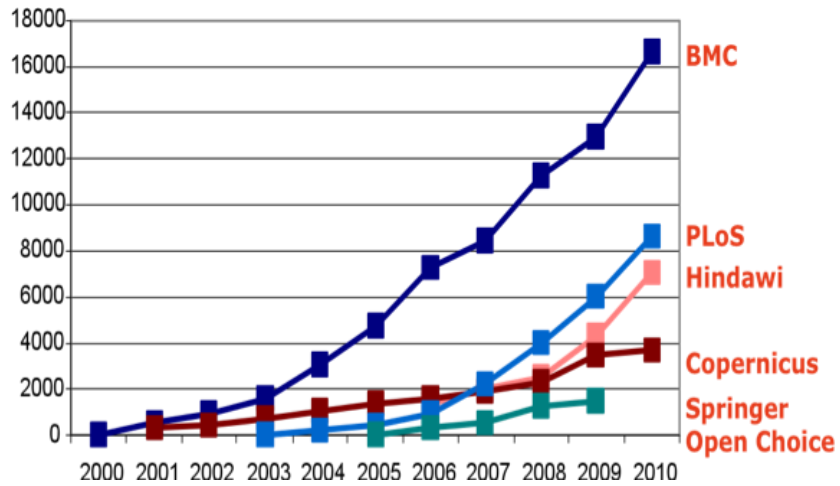
The first two were born as digital-only, OA publishers, while Hindawi was originally an SB publisher. OA publishers are newer organisations than the traditional academic publishers. Their characteristics are described in Table 6.

Table 6 Open Access publishers

	PLOS	Hindawi	BioMedCentral
Number of journals	7	265	207
Disciplines	Science and Medicine	Agriculture, Biology, Chemistry, Computer Science, Dentistry, Engineering, Environmental Sciences, Geosciences, Material Science, Mathematics, Medicine, Neuroscience, Nursing, Pharmacological Sciences, Physics, Social Sciences, Veterinary Sciences	Biochemistry, Bioinformatics, Biotechnology, Ecology, Genomics, Medicine, Public Health, Veterinary Sciences, Zoology
Year of start up	2003	2008 (as Gold OA)	2000

Figure 1 shows the evolution of the number of articles published per year.

Figure 1 – Evolution of the number of articles published by OA journals



Source: Patterson 2011

Institutions and firms running repositories vary significantly in terms of institutional scope (national or institutional versus discipline), institutional characteristics (institution, unit of a governmental entity, not-for-profit or commercial firms), type of material archived (citations versus full text), level of certification of the material archived (published versus unpublished material), services offered to the research community, involvement in the actual archiving of material, make/buy decisions concerning the platform, speed of growth of the repository,

direct access to authors. Table 1 in Annex 2 on page 72 lists the characteristics of the institutions taken into consideration.

With the help of the PEER project manager, interviews were arranged with representatives of the organizations highlighted in bold in Table 5. A PLoS representative was interviewed as an example of a gold OA publisher on a simplified version of the questionnaire; Università degli Studi di Milano institutional archive was added, due to the possibility offered to the research team to analyse the case; CSIC was suggested as a case to be added to the sample and agreed to be interviewed. Both repositories therefore were analysed using the same grid of analysis. These three extra cases are marked in italics in Table 5.

One face-to-face interview with at least one member of the organization was set up, either represented in PEER or referred to by PEER. It has been agreed by PEER and by the research team that access to individual actors would be possible via the contact person, who typically was not directly responsible for the management of the platform or the repository, nor part of the administrative and control department. This decision made it necessary for the research team to contextualise as much as possible information on each individual case, in order to gain sufficient confidence in the understanding of the relationship between costs presented by individual players and their drivers. Moreover, as costs reported by interviewees may suffer from differences among players in cost allocation criteria that could not be easily detected by the research team, the first interview was focused on the identification of the key information and on the level of granularity that made it possible for the interviewee to gather information internally.

A separate questionnaire was created for publishers and for institutions managing repositories. Individual respondents from the two types of organization were met separately during a face-to-face session or small group interviews and then interviewed during a series of telephone conversations and via email to gather additional information or to clarify unclear or underemphasised aspects. Tables in the annex show the areas investigated.

As far as publishers were concerned, in order to assess the costs of archived content it was necessary to determine costs associated with:

- Content certification. This made it necessary to understand how peer review is managed by the journals. The portfolio of journals in the sample includes:
    - Journals owned by the publisher and journals belonging to other institutions (for instance societies) outsourcing article publication to publishers.
    - Journals with different percentages of rejection rates, a different number of reviewers involved, a different number of rounds of revision.
- Content publication. Costs taken into consideration referred to formatting, cross referencing, metadata, typesetting, editing, translation into HTML and uploading to the platform hosting the content.
- Content management and preservation. This cost category includes depreciation of platform set up or cost of platform licensing, yearly maintenance costs (software and hardware), personnel costs associated with platform management.

For all cost categories highlighted above, interviews were aimed at determining whether different activities were performed in-house or outsourced. In the first case, personnel costs were calculated by determining FTE staff and the average cost provided by the company. In the second case, total costs billed by external suppliers were looked for. Cost drivers changed for different cost categories. For instance, costs associated with content certification as well as distribution are driven by the number of journals (typically rejection rates are calculated at the journal level). Cost per article is inferred by determining the average number of articles per issue and the average number of issues per year. On the other hand, costs per article associated with archiving and preservation are not only influenced by the number of articles archived and the yearly average intake of new articles, but also by the mix of digital documents archived. For gold OA publishers, the number of articles was considered a key cost driver.

Interviews with repositories were also aimed at determining the cost of archived certified material. Background information included date of set up of the repository, size of the repository in terms of number of certified articles and other types of material available, nature of references stored, number of researchers employed (for institutional repositories). In this case, cost categories differ in part and refer to:

- Content uploading to repository. Different channels to upload content were explored, namely self-archiving by authors, uploading by staff, metadata provided by publishers, harvesting.

- Content archiving and preservation. Costs in this category include set up costs or fees associated with platform management, software development and maintenance costs, long-term preservation.

- Marketing of repository and support to authors. These costs include presentations to researchers and support to authors during content self-archiving.

When internal staff were utilised to perform tasks, FTE and relative cost has been calculated; otherwise, the cost of the service outsourced was looked for.

Cost of compliance with PEER for publishers and repositories was determined by looking at the cost of dedicated staff associated with the project, the investment needed in software and the maintenance costs.

Table 2 in annex 1, page 70 highlights the information available to the research team on the publishers and Table 2 in annex 2 page 73 on the repositories interviewed.

Information may consist of hard data (EUR, number of people) or qualitative responses to a specific question ("We do not check for text searchability on our repository, as the responsibility of the activity is delegated to participating libraries"). Hard data in turn may be actual numbers (as in the case of externalised costs), estimates ("Our editorial staff works approximately 35% of its time coordinating peer review; we do not have any dedicated person to support authors in the self-archiving process, but we could say that 5% of our librarian time is devoted to support researchers in the self-archiving process") or standard costs provided by the administrative office ("The average cost for the editorial staff in our organization is 50.000 EUR per year").

Whenever possible, the research team sought to identify direct costs associated with specific activities. This was easy for outsourced costs. For instance, it was possible to determine the cost for externalised services for metadata and copyediting when these activities were performed by third parties. For internal costs, whenever possible, FTEs were identified and standard costs applied to determine salaries. Company overheads were not calculated, nor depreciation of assets taken into account.

As far as publishers' content management and distribution platforms are concerned, estimates were made on the set up investment, for publishers set up in recent years. For more established players, the most recent investment was quantified. The costs of outsourced services, of yearly maintenance costs, and FTE involved with platform management were also calculated. Similarly, in the case of institutions running repositories, the set up cost and maintenance costs of repositories were calculated, or the service cost to third-party platforms, together with FTE and maintenance costs.

Limitations of the study

The way this study has been designed has made it possible to examine the issue of the ecology of scholarly publishing associated with the growth in the diffusion of OA journals and repositories from a bottom-up perspective. By analysing individual publishers and organizational units managing repositories and focusing on the costs associated with research registration, certification and awareness, it has been possible to gain a sense of the actual costs incurred by different players under different business models and of different sizes and to appreciate the variety of conditions in which costs are generated. In this respect, this study is unique.

This method allows the researcher to gain insights into the specific issues addressed by individual players in their quest for sustainability and is particularly useful in these times of industry transformation. However, a few important limitations should be highlighted:

- As for any case based analysis, it can be argued that the results can only be generalised in part. The sample of publishers analysed includes diverse players in terms of size, scope, age and business model, but some types of players (for instance small gold OA publishers) are not represented. An extensive literature analysis was performed as a way to compare findings with published information.
- Some of the repositories analysed have been set up recently and they might not yet be operating smoothly. It may be that their impact has been underestimated, as the ramp up phase has not yet been completed.
- Case analysis is a relatively intrusive method of research, as it requires access to a large amount of information; the number of people interviewed for each organization in this study was very limited, as it was agreed that the research team would interface with one representative of the organization who would in turn gather the necessary information within the organization. To overcome this problem, the research team has tried to gather as much background information as possible from published sources on the individual organizations analysed.

- Information gathered through interviews is both quantitative and qualitative. To reduce bias, interviews were carried out by a very limited group of researchers and always in pairs.
- In commercial companies, information on cost structure is sensitive. The research team tried its best to balance the need to inform and document with respect for privacy.
- Cost assessment is influenced by organizational characteristics within each institution (make/buy decisions; task/cost responsibility allocation among units) and by the accounting and control criteria and practices followed by individual organizations, particularly as far as allocation of fixed costs is concerned. To reduce the impact of the variety of methods used to assess costs, the research team followed a direct costing approach, quantifying first costs that could be directly allocated to a specific organization and considering fixed costs in bulk.
- This approach has made it necessary to focus on some cost categories and to ignore others. For instance, the impact of overheads on overall average cost per article published has not been calculated.
- In both journals and repositories, some costs are either transferred to third parties (for instance peer review to external reviewers, or upload of metadata to authors or participating libraries) or are very hard to identify within a broader cost centre (as it might be a technical cost within an IT budget). Therefore some cost estimates might be inaccurate.
- In spite of their relevance, some costs simply could not be estimated. For instance, the quality, the value and the impact of a repository is significantly influenced by the choices made about collection policies or the effort made to ensure thorough content acquisition. As no exchange of goods or services takes place in the formation of these costs, they are simply overlooked.

4. Results

Publishers cost structure

Cost of published research outcomes for SB publishers in the sample results from taking into consideration the cost of content certification, content publication and published content management and preservation.

Cost of content certification

Reputation is a critical source of competitive advantage in scholarly publishing. Robustness of selection and the involvement of prestigious reviewers drives reputation. At the same time, peer review is a costly activity that can be standardized only marginally. Even if it is outsourced – and rarely remunerated – the publisher still has to bear the cost of managing peer review. Such costs correlate with the rejection rate of the journals, to the number of reviewers per manuscript and to the number of rounds of review. In some instances, dedicated editors in-house pre-select incoming manuscripts, thus reducing the number of those that go through peer review. This reduces the costs of finding reviewers and of managing the review process, but increases internal costs. In order to guarantee reputation on the one hand and cost control on the other, most publishers will have in their portfolio a group of journals with high rejection rates, reputation and impact and a significant group of more accessible journals.

Not all journals in a publisher's portfolio are owned by the publisher. In some cases, peer review is organised and managed outside the publishing house and publishers don't know the cost of content certification, or the rejection rate, but simply provide publication and distribution services.

Average cost of content certification per article published for the publishers considered is around 250 USD. The cost includes only salary costs and external fees paid for organizing and managing peer review. Results are lower than those reported by Houghton (2009) based on Tenopir and King (2000). It has to be noted though that only costs directly associated with this activity have been taken into consideration. Not surprisingly, the effect of content certification on the total cost of archived articles varies significantly among publishers in the sample.

Table 7 presents the case of four of the publishers in the sample, highlighting the different mix of journals published. Publishers A and C only publish journals owned by the company and for which peer review is organized and managed in-house. The two organizations differ in the average rejection rate of the journals published. Only a portion of the total articles published in one year by publisher B or D undergoes a peer review process managed by the publishers' editors; moreover, the mix of titles in the publisher's portfolio shows a different spread of titles with lower rejection rates and therefore a lower impact on certification costs.

Table 7 Journal distribution by rejection rate

Publisher	A	B	C	D
High (>80%)	12 (37,5%)	3 (7,1%)	6 (85,7%)	1- 2%
Medium (60-80%)	12 (37,5%)	34 (91,9%)	1 (14,3%)	15 - 20%
Low (50-60%)	8 (25%)			80%
Total in-house peer reviewed journals	32	37	7	1.800
Total journals (%)	100%	55%	100%	75%

Direct interviews with publishers show that internal costs (mainly FTE staff to coordinate peer review and editorial activity) account for 70 to 80% of editorial costs; no sign of significant economies of scale occur at editorial level, except for submission tracking.

Another element affecting cost of content certification is related to the complexity of the review process. Table 8 shows the number of submissions managed by editors, the number of articles accepted for publication, the number of reviewers involved and the number of reviews per reviewer. It is not uncommon that editors within the publishing company will pre-select articles to undergo the review process, in order to reduce the complexity and the cost (in economic or in organizational terms) of content certification.

Table 8 Complexity of content certification process; approximate number of articles

Publisher	Submissions	Accepted articles per year	Accepted articles (%)	Reviewers	Reviews
1	3.500	650	18%	2 to 4	2 or 3
2	21.000	7.000	33%	2	2
3	600.000	280.000	47%	2 or 3	2

Cost of content publication

Digital content publication involves many labour-intensive activities, including formatting, editing, typesetting. Cost of publishing (including metadata) ranges between 170 to over 400 USD per article, higher than the reported 200 USD by Houghton (2009). Management of publishing activities is heavily outsourced to specialized firms, mostly located in south-east Asia; bigger publishing groups have equity control over specialized firms, while others have commercial agreements with different pricing structures. Publishers of journals written in English enjoy a cost advantage over colleagues publishing in different languages in this respect. Moreover, in our sample there is clearly an economic advantage in externalising

content publication.

The cost of publication per article was calculated by asking interviewees to report on the number of FTE, average salary, total external costs associated with the total or partial subcontracting of this activity. Some of these costs were available per journal; in this case, cost per article was calculated by dividing the cost per journal by the average number of articles. Journals characterised by a high number of articles per issue and a high number of issues may therefore appear with a lower cost of publication per article.

Cost of content management

Table 9 compares the size of the catalogues of the SB and OA publishers in our sample

Table 9 - Catalogue size

	Number of articles in catalogue 2010	Number of articles published 2010	Journals 2010
BioMed Central	62.000	16.500	215
BMJ Group	824.183	51.728	32
Cambridge (CUP)	700.000	45.000	239
EDP Sciences	146.500	7.000	47
Elsevier	11.000.000	300.000	2.392
Hindawi	14.000	7.500	265
IOP Publishing	400.000	30.000	67
PLOS	15.245	8.000	7
Sage Publications	757.000	30.500	560
Springer	4.000.000	140.000	2.000
Taylor & Francis	4.000.000		1.500
Wiley-Blackwell	4.000.000	120.000	1.500

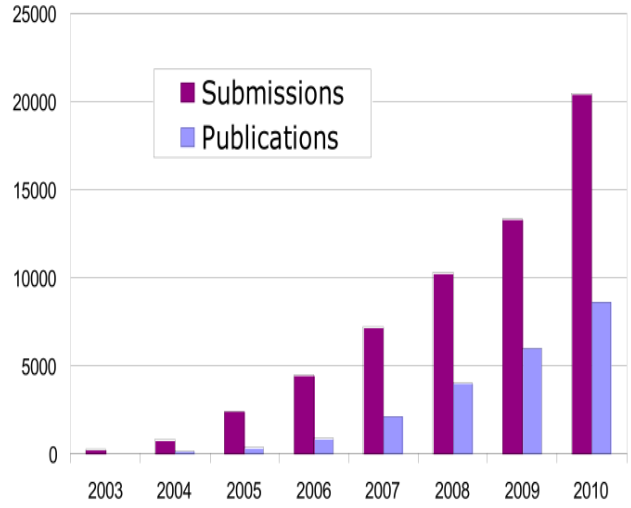
For a publisher the choice of the platform is strategic. Platform characteristics affect article layout, the services provided to authors and users and the possibilities for interfacing with internal and external IT infrastructures. Platform configuration therefore affects functionality and the positioning of the publisher. It is a choice characterized by high resilience for technological, economic and organizational reasons; for a publisher and a repository, there are significant switching costs in shifting from one platform to another. Specific studies have been conducted to compare different solutions (see for instance Crow 2004; Repository Support Project 2010).

Content management and user-interface platforms may be proprietary, or outsourced. Publishers may develop proprietary platforms for content management and distribution - as is the case for Elsevier, Wiley-Blackwell and PLoS – or outsource hosting and dissemination

services to organizations like MetaPress or Highwire, as is the case for Springer or Sage. In some cases, journals belonging to the same group may be hosted on different platforms, because the previously belonged to a different publisher.

The decision to internalise the development of a platform pushes the publisher to increase the number of journals and articles published, and possibly to increase the scope of their activity, so as to spread a significant fixed cost over a broader number of documents. In this case, the cost of platform management is subject to significant economies of scale. Consider the case of PLoS. Between 2003 and 2007 the publisher published six top-quality journals, successfully aiming at reputation and brand building via selective peer review and highly qualified editorial boards. Two of them publish on average 10% of the submitted manuscripts. Figure 2 shows the evolution of submitted and published articles.

Figure 2 – Submitted and published articles - PLoS



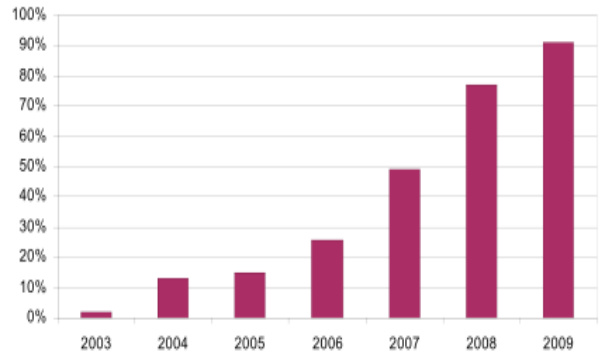
Source: Patterson 2011

In December 2006, PLoS One was added, characterised by a broader scope and a higher acceptance rate. The introduction of PLoS One has made possible important innovations in the services provided to authors and readers and set the stage for reaching operational sustainability. Speed of processing and richness of information around and about each article make it possible to have a broader understanding than with a simple impact factor of the contribution of an article to the research community[33]. Figure 3 shows the evolution in the percentage of operating costs covered by OA revenues. Operating costs include depreciation for updates of the platform, but do not cover the initial investment to set up the platform, which was covered by a series of donors[34].

[33] https://indico.cern.ch/contributionDisplay.py?sessionId=8&contribId=20&confId=103325

[34] 9 million USD as reported by King and Alvarado Albertorio (2008).

Figure 3 – Evolution of the percentage of operating costs covered by operating revenues



Source: Patterson 2011

Platforms may be developed around open source software, as is the case of Ambra developed by PLoS; the advantage of open source software is that development costs and subscription costs are spread. Moreover, interoperability and collaborative efforts are facilitated.

Internalisation of platform development stimulates growth in the number of journals and articles published; this in turn leads to the internalisation of a significant amount of information on user behaviour and potentially to developing targeted services, including advertising. Publishers with their own platform may decide to reduce the costs associated with platform management by licensing the platform to other publishers, possibly in different disciplines so as to minimize competition.

Platform costs can be roughly divided into two categories:

- Set up investments and major improvements.
- Annual maintenance costs.

As expected, for publishers internalising platform and content management the most difficult item to quantify is the investment for content archiving within the overall IT budget. Claimed overall investments ranged from less than 1 million USD to tens of millions for one big commercial publisher.

Interviews with publishers on the size, characteristics and evolution of their proprietary platforms and on set up costs gave results of recent investments for significant technological upgrades of the order of hundreds of thousands of dollars. Technological upgrades occur typically every other year. Platform set up costs are obviously driven by their configuration and by the technological possibilities available when they were first designed. There are costs associated with protection and user friendliness and openness.

Maintenance costs are somewhat easier to account for and discrepancies among companies in the sample are much lower. Publishers interviewed report costs ranging from USD 170k (for platforms that are de facto content management systems) to 400k for highly sophisticated architectures. Companies externalising platform management, typically pay an annual fee partially related to the number of documents hosted, in part to the types of services requested.

Compliance with PEER

The majority of publishers interviewed could not provide detailed information on the cost of compliance with PEER, either because costs were, all in all, perceived as marginal or because they were perceived as sunk costs, and therefore difficult to quantify. The two main cost categories mentioned were software development/adaptation and human resources (meetings, planning, project team). One publisher claimed a total cost of over 100.000 USD, but the other publishers who did provide data reported two days per month of personnel costs and expenses in the range of 10-15.000 USD. The structure of the platform, criteria used to allocate costs and the inclusion of sunk costs may explain the differences in the responses obtained.

Repository cost structures

Cost of content uploading

Table 10 shows the characteristics of the repositories analysed in this study at the end of 2010[35].

Table 10 – Content composition of repositories analysed

	Max Planck Society (eDoc)	**Inria/CNRS (HAL)**	**CSIC**	**Göttingen University (GoeScholar)**	**Gesis**
Date of set up	2010	2005	2007	2009	2008
Material archived					
Number of items archived	153.905	20.252	25.618	2.300	6.818
New items during 2010*	16.902	24.426	Growing	approx.800	3.000
Format of full text	Various	Pdf	Pdf	Pdf	Pdf
Full text searchability	not checked by Max Planck, but by research institutions	Yes	Yes	yes	Yes
Nature of material					
Bibliographic reference (%)	85%	28%			

[35] Data on GoeScholar refer to November 2011

Reference and link (%)		1%	20%		20%
Full text (%)	18,16%**	71%	80%	100%	80%
Total number of full text documents	27.641	14.309	20.000	2.300	5.450
Type of full text material					
Unpublished material (%)	60%	57%	20%	0,5%	30%
Published material and stage 2 (%)	40%	42%	80%	99,5%	70%
Feeding the archive					
Self-archived by authors (%)	n.a.	Yes	18%	25%	Yes
References inserted by staff (%)	95%	Yes	37% ***	25%	Yes
References provided by publishers (%)	1%	No		33%	50%
References gathered from various sources (%)	To be implemented PubMan	link with PubMed and ArXiv	45%***	17%	20%

*= The number of new references per year may vary significantly. Max Plank Society shows this evolution:

2003	2004	2005	2006	2007	2008	2009	2010
4.799	6.853	9.658	41.180	14.985	39.345	16.902	18.064

The growth in the number of items archived may change quite significantly, due to the fact that many repositories are very young. For instance, Gesis was set up on January 2008 and as of September 2011 has archived approximately 20.000 documents.

**= The percentage refers to full text but not necessarily publicly available full text. Reference with publicly available full text: 11.682 (7,67%).

***= 37% refers to the staff at CSIC technical centre, whereas 45% is uploaded by librarians at different research centres

All repositories analysed are developments of publicly funded institutions to support high-level research at a national level or in specific areas. Setting up a repository therefore is usually part of their mandatory functions. The repositories mentioned are relatively recent developments, with MaxPlanck eDoc being the first in the sample to be set up. Given the limited size of the sample and the short existence of some of the cases analysed, the evolution of content uploaded is shown only for the two biggest repositories considered (see Table 11)

Table 11 – Evolution of references in repositories

	Max Planck Society	Inria/CNRS
2002	441	
2003	4.799	
2004	6.853	
2005	9.658	1.083
2006	41.180	11.262
2007	14.985	14.151
2008	39.345	17.426
2009	16.902	24.426
2010	18.064	

All repositories in the sample provide documents in full text searchable Portable Document Format (PDF). In the case of the Max Planck repository, the responsibility for upload rests on the individual institutes and therefore the Max Planck Society does not have the responsibility for verifying searchability. The nature and composition of repositories vary, with all repositories, except that of Max Planck's eDoc offering for the most part full text material. The percentage of certified research outcome varies quite significantly, with eDoc and HAL repositories allowing for a higher percentage of non-published/non-refereed materials. Max Planck and CSIC differ from the rest of the sample in the way they organize content collection and processing; in both cases, the greatest effort in processing material and interfacing with individual groups of affiliated researchers is managed by the libraries at the local level, with the repositories at the central level acting as a coordinating and support player.

All interviewees had difficulties in assessing the relative importance of different players in the deposit process of scientific contributions into the repository. Librarians seem to play a key role in processing information and performing quality control on the material hosted by the repository, although not necessarily at the central level, as has been mentioned for Max Planck Society and CSIC. The outsourcing of processing responsibilities at the local level reduces the impact of processing costs on the repository's cost structure; as can be seen from Table 11, human resources requirements for repositories are limited. Because of the impact of externalities, the reported cost of processing per document for repositories is 10 EUR maximum per reference archived and 18 EUR maximum per full text document, and 43 EUR maximum per journal article. The cost includes metadata creation.

At the same time, repository managers acknowledge that the process of delegating to local

libraries part of the processing may delay deposit in the repository and therefore slow the repository's growth. Harvesting and mandates are effective ways to speed up the feeding of the repository, but the possibility for repository managers to actually enforce such mandates is limited.

A repository's growth rate is considered an important element in determining its effectiveness and its ability to play a role in research dissemination. The issue of growth and critical mass is important for repositories as well as for publishers; as a reference, Table 12 compares the data on the repositories analysed with other (particularly subject-based ones) that have succeeded in reaching a solid reputation and developed steady growth. A rich content base is viewed as a prerequisite for user attraction[36] and therefore for articles downloaded. The literature already acknowledges that the ramp up phase for journals is six years; data in Figure 1 for OA journals confirms it. As most repositories are younger than six years, and as no calculations have been made on cost per download for journals, it seemed appropriate to calculate costs per upload only.

Interoperability is viewed as an important ingredient to allow a more effective collaboration among OA journals and repositories.

Table 12 -- Size of repositories

	References	**Full text**
ArXiv	673.024	673.024
CSIC repository	26.230	21.508
SSOAR	6.818	5.450
GoeScholar	2.300	2.300
HAL	27.382	14.309
eDoc	153.905	25.968
PubMedCentral	24 million (PubMed)	1,8 million
SSRN	270.109	270.109

[36] PubMedCentral holds OA articles deriving either from the publication in an OA journal or as author manuscripts complying with mandates and is searchable via PubMed.

Repository set up and maintenance costs

Determining the actual cost of the platform and its maintenance costs proved to be very difficult for two main reasons: investment in platform set-up, and costs in software upgrade and repository maintenance are sunk costs, that is to say they are either not monetized or are items of cost centres within the institutions that are not controlled for separately. Moreover, the organizational unit within the institution in charge of repository management typically has a budget that covers only some aspects associated with repository management. Repository set up and technical upgrades are in general under the responsibility of the IT department. Technical costs associated with repository management are calculated in terms of FTE allocated to the organizational unit in charge of repository management.

No data could be extracted to determine the set up cost of the repository and only in one case was it possible to extrapolate external costs for software development. Services include hosting, maintenance support, usage reporting; sunk costs are related to the need to personalize or interface the platform. Repositories interviewed reported costs in the range of USD 60k for internal software development (typically personalization of platform) and hardware.

On the other hand, all repositories interviewed were able to calculate the number and cost of FTE of technical staff devoted to the repository. As has been stated in respect of publishers, repository set up costs and maintenance are heavily scale driven, thus – notwithstanding the major limitations highlighted – the cost per item in the repository tends to be lower the bigger the repository. It ranged between 2 and 50 EUR per reference and between 2,5 and 53,2 per full text journal article.

Support and author involvement

The last category of costs considered deals with collecting references and offering support to authors self-archiving material in Green OA. The costs of these activities are for the vast majority associated with personnel costs for personal or virtual assistance, and for promotion of OA within the institution. Interviewees reported between 1 and 3 FTE for each institution.

Involvement in PEER

Managers of repositories interviewed on the involvement of their institution in the PEER project showed very different levels of commitment and cost. Personnel involvement ranged from between 2 hours per week to 3 FTE per year. Set up investments ranged between 4.000 and 10.000 EUR; one institution allocates 1,5 technical FTE to software development.

5. Discussion and interpretation of findings

Comparison of cost structures and cost drivers

The empirical research described in the previous paragraphs on the costs associated with research certification, publication and digital management by a sample of journal publishers and repositories highlights the following elements:

- The average cost of content certification per article published for the publishers considered is around 250 USD; the cost includes only salary costs and external fees paid for organizing and managing peer review. No sign of significant economies of scale occur at the editorial level, except with respect to submission tracking.
The incidence of content certification on the total cost of archived articles varies significantly among publishers in the sample, and it depends on the journal rejection rate and complexity and length of the review process. To reduce the impact of costs of content certification on total costs, publishers include in their portfolio journals with differing rejection rates.

- The average cost of publishing (including metadata) ranges from 170 to over 400 USD per article, and is influenced by make/buy decisions and by journal size; publishers publishing journals in English are favoured by outsourcing in low-wage cost countries.

- In order to make content accessible, it has to be managed via a digital platform allowing content management, storage and accessibility. Costs associated with digital platforms vary significantly, depending on whether the platform is proprietary or based on open source software, on the age and characteristics of the platform, on the number of articles and documents stored and on the complexity of the platform in terms of services offered to readers and authors. Therefore, the effect of platform investment on article costs is extremely hard to calculate and shows a high variance across the publishers interviewed. Moreover, in the case of some repositories and OA publishers, platform set up investments have been covered by a grant, so that cost structure does not include initial investment repayment. Open source platforms allow for shared costs of upgrades and easier interoperability, but the research team was unable to quantify cost differentials with proprietary platforms in this respect. Maintenance costs are somewhat easier to account for and discrepancies among companies in the sample are much lower. Yearly maintenance costs range from USD 170k to 400k. The effect of these expenses on the average cost per article depends on the size of the publisher in terms of the number of journals it publishes, the number of articles per journals and the complexity of the platform in terms of services offered to authors and users.

- The availability of open source publishing platforms, as reported by Edgars and Willinsky (2010), allows for a drastic reduction of certification, publication and platform management costs, thus reducing barriers to self-organization of groups of researchers to publish journals and favouring the start up of new journals. One of the outcomes of the OA movement has been the opening up of publication opportunities to groups of researchers and learned societies on a large scale. A significant proportion of these costs is transferred to groups of researchers who provide voluntary labour associated not only to content certification, but also to all aspects related to management of a journal. Given the low-scale effect on certification and publication costs, it is possible that several new OA journals heavily relying on volunteer work publish a very limited number of papers.

- Management of repositories is cost effective from an operational point of view and in line with the mandatory functions of the institutions analysed. The use of open source software and the effort to participate in collaborative projects allow for very limited direct costs. However, the potential impact on scholarly research is related to the availability of a critical mass of well organized, visible and easily accessible research outcomes. Speed of growth in the number of articles offered and richness of services to the scholarly community, such as usage and citation statistics, rankings, is therefore going to become increasingly important as the amount of openly accessible content increases.

- Some repositories are characterised by a high level of sunk costs. For instance, in the case of organizations coordinating several research institutions, costs of uploading articles and adding metadata are transferred to participating libraries; moreover, costs of software maintenance and upgrades are hard to identify as they are sunk in overall IT budgets. The potential for the organizational unit managing the institutional repository to effectively enforce mandates is limited and is often related to time-consuming internal diplomacy, which is not accounted for. The time and energy required to a researcher to self – archive is limited, but repositories may encounter organizational difficulties in ensuring that all researchers within the organization correctly interpret the opportunities associated with the upload of articles into repositories and actively contribute to its growth.

- From the aggregate point of view of research registration, certification awareness and preservation, individual organizations define their scope based on a series of make/buy choices for the different activities performed. Make/buy choices affect cost structure, which in turn determines the extent to which sustainability of a specific business model is affected by size, i.e. by the overall number of articles managed. Make/buy choices refer to the following aspects in the cost categories considered:

As far as publishers are concerned, make/buy decisions affect all phases analysed.

- Content certification: peer review may be organised by a journal editor within the publishing house or externalised.

- Content publication. Activities associated with this cost category (copyediting, typesetting, and proofreading) may be performed by the publisher's staff, outsourced to third parties (often located in low-wage countries) or be performed by the author.

- Content management, hosting and distribution. Publishers may manage content in a proprietary platform or pay a fee to use third-party platforms (e.g. HighWire, MetaPress). Publishers with proprietary platforms may decide to license the use of the platform.

For repositories, make/buy decisions affecting cost structure refer to the following activities, associated with different cost categories:

- Content uploading to repositories. Institutions managing repositories may use staff to upload content, harvest, outsource the activity to publishers paying a fee or to authors. Make/buy choices affect not only the cost of the activity, but also the speed of growth of the repository and quality control.

- Content archiving. Repositories may have developed their own open source platform, or pay a fee to adapt existing software or use a third-party platform. In this case, maintenance and costs related to platform upgrades are included in the fee paid to the organization in charge of platform development and management.

The OA movement has been successful in challenging and removing some of the market imperfections of scholarly publishing and to offer scholars and institutions alternative ways to publish and access quality research - by making it possible to treat content publication and dissemination separately and by maximising accessibility to often publicly funded research. Moreover, the presence on the market of different publishing options puts pressure on prices, quality of offering and efficiency. The birth and diffusion of open platforms allowing for low-cost publication and the diffusion of OA journals and repositories has made it possible for scholars to publish and access journals that might otherwise be cut by pressure on library budgets.

The not-for-profit nature of some OA publishers and repositories made it possible for them to gain access to grants to cover costs to set up the digital platform on which articles are stored. Now that some of them succeeded – individually or collectively - in creating a critical mass of openly accessible research outcomes, the issue of their sustainability is not just a matter of the efficiency of individual organizations but also of the interplay of scholars, research institutions, publishers and funding agencies.

The ecology of scholarly publishing: competition in multisided markets

As the number of OA journals and repositories increase and become familiar to a broader number of scholars, the issue of sustainability and competition for resources and reputation will increasingly affect both OA and SB players as part of the same competitive environment. Six elements seem to the research team challenging in the near future:

- The growth in the supply of documents. There is evidence of a steady growth of research outputs, driven by a variety of factors, that are being researched[37]: they include globalization of research and an increase in the number of researchers engaged in publishing, decreased publication times, a rise in the number of articles per journal; the development of novel formats taking advantage of web 2.0 (newsletters, blogs, commentaries); the availability of intermediate outputs before the certified published version; the variety of "new content" that is worth organizing, storing, sharing to make research more effective and efficient (protocols, clinical studies, datasets). Both SB and OA publishers, as well as groups of researchers and repositories, contributed and are contributing to transform this content into organised material that could be accessed and used and possibly become a source of competitive advantage for whoever succeeds in effectively transforming it into valuable service to the research community.

- The fragmentation of industry practices and interdependence among players. Besides new models for content publication and distribution, an array of start-ups is emerging to deal with

[37] Systematic literature analysis on these aspects was omitted in this study

the selection, validation and organization of content besides articles; other players focus on researchers' behaviours and patterns, to identify useful services. This group of "cultural entrepreneurs" (partially private, partially not-for-profit), once fully established is likely to contribute to the transformation of the industry towards a higher importance attributed to information services in addition to certified content. Of particular interest are growing instances of successful public private partnerships:

> One interesting area of cooperation between publishers and libraries concerns content preservation. JSTOR is a consolidated example of a collaborative project started with public funding that has been able to grow into a self-sustaining business model offering services to libraries, institutions and publishers for content preservation of journals. Another example of cooperation is CLOCKSS, a project associated with preserving discontinued digital journals, offering an important service to the scientific community at a reasonable cost.
> Other possible initiatives are associated with research institutes or groups of researchers developing coalitions (such as SCOAP3) to negotiate deals to publish research outcomes and to guarantee OA at the same time as quality of certification and reasonable remuneration for entrepreneurial effort.
> The still largely unexplored area of shared data opens significant opportunities for innovation and for cooperation, as new technical as well as organizational, editorial, legal and economic solutions need to be developed[38].

- The incentives set up by a variety of research institutions: Funding agencies are increasingly putting pressure on scholars to make their research publicly available, thus favouring de facto OA journals; at the same time, incentives to researchers by research institutions to publish on highly visible journals may hinder the diffusion of OA journals, making the scenario of a substitution of SB to OA unlikely[39]. The interplay of incentives of different nature affects OA diffusion across disciplines, while in general putting pressure on smaller and less visible journals, both SB and OA.

- The economy of attention: In spite of the increased effectiveness in organizing and making content available, the number of articles a researcher will read and cite is limited and competition for researcher's attention is fierce. Efforts to concentrate large amounts of content into publicly available and freely accessible repositories, and the coordinated effort of institutions and funding agencies on OA policies have and will reduce commercial publisher's competitive advantage that rests on artificially limited and scarce access to content by means of copyright restrictions. At the same time, freely accessible content is likely to be a necessary but not sufficient condition for attracting readers. Both repositories and journals will therefore be pushed to develop services for authors and readers.

[38] See for instance Riding the Wave: How Europe Can Gain From The Rising Tide of Scientific Data, EU working group, 6 Oct. 2010 http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf

[39] On the benefits of a more aggressive mandatory policy see for instance Houghton et al. 2010. Commenting abut the sustainability of OA journals, Suber (2009a) acknowledges the different level of maturity of different business models: "if comparatively little is spent today on OA journals, that says more about the history of journals (in which TA journals arrived long before OA journals) than about the sustainability of OA journals" (point 8).

- The preservation of memory: As the amount and variety of content produced increase, the problem of guaranteeing appropriate preservation of published research and of other relevant material in digital form becomes increasingly important. Specific resources need to be devoted in order to provide effective, efficient and secure storage and accessibility of content over time.

- The overall financial crisis and a generalised strain on resources: The current economic crisis faced by European countries will put a pressure on available resources at all levels, increasing the competition among research groups to access resources and the competition between journals to attract research outputs. SB journals need to more carefully tailor their deals with libraries, which in turn are increasingly seeking ways to maximize the potential of the deals, while at the same time reducing lock-in effects associated with deals. OA journals are seeking agreements with research institutions to develop suitable schemes to cover article-processing fees. It is likely that funding agencies will pay more attention to costs and benefits associated with alternative resource allocation, and pressure on journals and repositories will increase to specifically address sustainability. Researchers are likely to devote more time and energy than in the past to fundraising for their studies. In their 2006 study for the European Commission, Dewatripont et al. recommended that education and research funding authorities should guarantee sufficient attention and resources to allow for competition and experimentation with different publishing business models, while also playing a crucial political role in shaping the evolution of the industry. Recent documents by the European Commission confirm the commitment to OA (COM(2010)245, 19.05.2010; COM(2010)546, 06.10.2010).

In this scenario, smaller players operating in budget tight disciplines are more likely to suffer, unless they are able to develop cooperation schemes around platforms. Dewatripont et al. (2006) present scientific publishing as a two-sided market (Rochet Tirole 2003), in which publishers act as intermediaries between authors and libraries and build their competitive advantage by bundling certification and awareness. The economics literature refers to multisided markets, (Rochet Tirole 2003) and defines competitive settings characterized by the presence of two distinct sides, whose ultimate benefit stems from interacting through a common platform (Dubini Giglia 2009). Products and services that bring together groups of users in two-sided settings provide infrastructure and rules to facilitate each groups' transactions and favour the loyalty of both sides to the platform. Platforms may be physical products (for instance videogames) or virtual markets (as in the case of credit cards) providing services (Eisenmann, Parker and Van Alstyne 2006). The distinctive feature of platforms is the need to simultaneously satisfy both sides of the market in order to prosper. In order to reach critical mass, platforms often treat one side as a profit centre and the other as a loss leader, or, at best, as financially neutral (Rochet, Tirole 2006). Therefore, the volume of transactions between end-users depends on the structure and not just the overall fees charged by the platform.

Organizations competing on two-sided markets may opt for different strategies to gain a foothold in both markets. In the case of SB models, resources gathered on the reader side, via subscriptions to libraries, are used to strengthen the offering to authors. On the other hand, OA models subsidize the reader side by gathering resources from the author side. When

content is freely available, as in the case of OA, libraries no longer cover publication costs directly. Rather, libraries – or funding agencies, or authors – directly finance a publication by covering article-processing charges or indirectly support OA by encouraging affiliated researchers to use OA platforms. The OA movement has therefore taken advantage of the possibilities offered by digitization to provide an alternative platform-based competitive model, in which the production side subsidizes the consumption side of the platform and not vice versa as in the traditional model. While profitability in SB models is driven by the number of consumers, e.g. subscribers, and by the number of pricing schemes that can be offered to them, the OA model builds its profitability by progressively enlarging potential revenues on the production side, for example by charging article-processing fees or institutional memberships.

Developing a successful platform-based strategy is not easy, as alternative business models may interfere. The research community (and the readership) appreciate the possibility of accessing research outcomes free of charge, and this is a point in favour of OA journals. However, there is still resistance to OA publishing (Dallmeier Tissen et al. 2011). A lack of research funding to cover publication costs, bias versus OA, career incentives set up by institutions related to publication in specific journals, no tradition of exchanging work in progress within the scientific community and fear of sunk costs are all elements hindering the diffusion of the OA movement. Moreover, when they are affiliated to research institutions, libraries act as a third payer, so many readers may still be unaware of the different nature of the two business models.

The interplay of the six issues described above and the progressive articulation of both OA and SB business models make the development of business models progressively more complicated and show the growing importance of specific platforms and specific players in operating simultaneously in several markets: authors, research institutions, funding agencies and readers.

Only a few players are currently able to attract large numbers of articles and readers and are therefore in the position to shape scholarly publishing:

- Among the open source repository software solutions, DSpace is the one used by the largest number of organizations and institutions, but other open source platforms, such as Ambra, are designed to provide better services to readers. Currently DSpace is used by over 1.000 repositories and publishers.

- Among the scholarly publishers, the Elsevier platform manages 11 million articles and has a remarkable presence among libraries; competitors such as Wiley-Blackwell, Taylor & Francis and Springer each manage approximately 4 million articles, with different strategies for SB and OA models. The ability of the main traditional publishers maintaining a grip on the enormous catalogue of already published research and the level of investment in virtually all aspects of content selection, publication and dissemination are strategic elements that will control the evolution of the industry. Moreover, they might nevertheless resist investing in community services around both content from their journals and OA content..

- HighWire is another interesting example of a platform that has successfully gathered 4 million articles, but with a different business model from that of

traditional publishers. It could engage in developing new services to publishers and libraries alike.

- By converging with PubMed, PubMedCentral is clearly a starting point for research, benefitting OA and SB journals. New alliances are developing between research institutions and funding agencies to set up an OA Journal[40].

This is not to deny the importance of other players or platforms (for instance OJS) in affecting the evolution of scholarly publishing, especially in specific disciplines. However, due to their size and influence on all actors involved with scholarly publication, the players identified are in a good position to directly or indirectly attract resources, attention, content and shape the evolution of scholarly publishing. The other players have to address the issue of reaching a significant size in a relatively small time. PLoS, BioMed Central and Hindawi have during the last decade been able to outperform other OA publishers in terms of reputation, service, efficiency and growth rate. Smaller OA journals might have to address the issue of critical mass; once big research communities are being served (some disciplines are intrinsically bigger in number of researchers and volumes of output produced), growth becomes harder to achieve and may require partnerships.

In other cases, for instance ArXiv, the business model is successful from the point of view of its ability to attract the research community, but is struggling from an economic point of view. The difficulty for ArXiv is to cover the estimated 2012 operating budget of 500.000 USD[41] and suggests that even prestigious and widely accessed repositories may not count on institutional financial resources alone.

[40] See for instance http://www.hhmi.org/news/schekman20110711.html
[41] http://arxiv.org/help/support/whitepaper

Bibliography

Quoted references

Armbruster, C., Romary, L. (2009) "Comparing Repository Types. Challenges and barriers for subject-based repositories, research repositories, national repository systems and institutional repositories in serving scholarly communication", International Journal of Digital Library Systems, Vol. 1, no 4, 2010, p. 61-73. Electronic copy available at: http://ssrn.com/abstract=1506905.

Ambruster C., A European Model for the Digital Publishing of Scientific Information?, SSRN, 2009. Available at http://ssrn.com/abstract=1106162

ARL Statistics http://www.arl.org/bm~doc/arlstat08.pdf

Barton, M.R., Walker, J. (2003) "Building a business plan for DSpace, MIT Libraries' digital institutional repository", Journal of Digital Information, Vol. 4, no 2.

Basefsky, S. (2009) "The End of Institutional Repositories and the Beginning of Social Academic Research Service: An Enhanced Role for Libraries". Electronic copy available at: http://www.llrx.com/features/socialacademicresearch.htm

Bailey, C.W. (2008) "Scholarly electronic publishing bibliography", 2007 Annual Edition.

Bailey, C.W. (2008) "Institutional Repositories, Tout de Suite", Digital Scholarship.

Baumol, W.J. (1972) "On Taxation and the control of externalities", The American Economic Review, Vol. 62, No. 3, Jun., 1972, 307-322.

Benkler, Y. (2002) "Coase's Penguin, or, Linux and the Nature of the Firm", Yale Law Journal, Vol. 112, no 3, p. 369–446. Preprint. Electronic copy available at: http://arxiv.org/ftp/cs/papers/0109/0109077.pdf

Bergstrom, T. (2001) "Free Labour for Costly Journals?", Journal of Economic Perspectives. Vol. 15, no 4, 2001, p. 183-198.

Björk, B-C. (2007) "A model of scientific communication as a global distributed information system", Information Research, Vol. 12, no 2, 2007.

Björk, B-C., Hedlund, T. (2009) "Two Scenarios for How Scholarly Publishers Could Change Their Business Model to Open Access", The Journal of Electronic Publishing, Vol. 12, no 1, 2009. Electronic copy available at: http://dx.doi.org/10.3998/3336451.0012.102

Björk, B-C., Welling, P., Laakso, M., Majlender, P., Hedlund T., Guonanson, G. (2010) "Open Access to the Scientific Journal Literature: Situation 2009", PLoS ONE 5(6): e11273.

doi:10.1371/journal.pone.0011273

BOAI, Budapest Open Archives Initiative, Read the initiative, 2002 http://www.soros.org/openaccess/read.shtml

Boldt, A. (2010) "Extending ArXiv.org to achieve open peer review and publishing", Journal of Scholarly Publishing, Vol. 42, No 2, 2011, p. 238-242. Electronic copy available at: http://arxiv.org/abs/1011.6590v1

Bot, M., Burgemeester, J., Roes, H. (1998) "The Cost of Publishing an Electronic Journal: A general model and a case study", D-Lib Magazine, November 1998. Electronic copy available at: http://www.dlib.org/dlib/november98/11roes.html

Bowen, W.G. (1996) "JSTOR and the economics of scholarly communication", The economics of information in the networked environment. Proceedings of the Conference challenging the marketplace solutions to problems in the economics of information, edited by M. A. Butler and B. R. Kingman. Washington DC: Association of Research Libraries, pp. 23–34.

Boyce, P.B. (1997) "The AAS Program of Electronic Publication" in Heck, A., Ed. Electronic Publishing for Physics and Astronomy. Astrophysics and Space Science Library, 224-55

Boyce, P. B. (1998) "Costs, archiving and the publishing process in electronic STM journals" in Against the Grain, Volume: 10, Issue: 6, 24-25.

Calcagnini, G., Saltari, E. (Eds) (2009) "The Economics of Imperfect Markets. The Effects of Market Imperfections on Economic Decision-Making" Heidelberg, Physica-Verlag GmbH & Co.

Carr, L., Harnad, S. (2005) "Keystroke Economy: A Study of the Time and Effort Involved in Self-Archiving" Technical Report, ECS, University of Southampton. Electronic copy available at: http://eprints.ecs.soton.ac.uk/10688/1/KeystrokeCosting%2Dpublicdraft1.pdf

Cavaleri, P., Keren, M., Ramello, G.B., Valli, V., "Publishing an e journal on a shoe string: is it a sustainable project?" Economic analysis and policy 39, 1 March 2009 89-100. Electronic copy available at: http://tsc.library.ubc.ca/index.php/lindsaywillson/article/viewFile/34/39

Chesbrough, H., Rosenbloom, R. (2002) "The role of the business model in capturing value from innovation: evidence from Xerox Corporation's technology spin-off companies", Industrial and Corporate Change, Vol.11, no 3, 2002, p. 529-555.

Chen, F.L., Wrynn P. Rieke, J.L. (2001) "Electronic journal access: how does it affect the print subscription price?", Bulletin of the Medical Library Association, Vol. 89 No.4, 2001,

pp.363-371.

Clarke, R. (2007) "The cost profiles of alternative approaches to journal publishing", First Monday, Vol. 12, no 12, 2007.

Cooney McQuath, S., Busch, S., Kahn, D. (2010) "Open access publishing: a viable solution for society publishers", Learned Publishing, Volume 23, Number 2, April 2010, 101-105. Electronic copy available at: http://docserver.ingentaconnect.com/deliver/connect/alpsp/09531513/v23n2/s4.pdf?expires=1317472867&id=64668191&titleid=885&accname=Guest+User&checksum=B794937D2D74BED1E4F70E435FEA8099.

Cornes R., Sandler T. (1986) "The theory of externalities, public goods, and club goods", Cambridge University Press, 1986.

Cox J. (2002) "Pricing Electronic Information: A Snapshot of New Serials Pricing Models", Serials Review, Elsevier, 2002.

Crow, R. (2004) "A Guide to Institutional Repository Software", Version 3.0, Open Society Institute. Electronic copy available at: http://www.soros.org/openaccess/software/

Cyzyk, M., Choudhury, S. (2008) "A survey and evaluation of open-source electronic publishing systems", Unpublished paper, Sheridan Libraries, Johns Hopkins University, Baltimore, MD. Electronic copy available at: https://wiki.library.jhu.edu/display/epubs/.

Dallmeier-Tiessen S. et al. (2011) "Highlights from the SOAP project survey: what scientists think about Open Access Publishing" ArXiv 1101.5260 Available at http://arxiv.org/ftp/arXiv/papers/1101/1101.5260.pdf

Darnton, R. (2009) "The Case for books: past, present and future", New York, Public Affairs.

David, P. (2003) "The Economic Logic of "Open Science" and the Balance between Private Property Rights and the Public Domain in Scientific Data and Information: A Primer" SiEPR Discussion Paper 02-30. Electronic copy available at: http://www.nap.edu/catalog.php?record_id=10785

David, P. (2004) "Can 'Open Science' be Protected from the Evolving Regime of IPR Protections?", Journal of Institutional and Theoretical Economics, Vol. 160, no 1, p. 9-34. Electronic copy available at: http://www-siepr.stanford.edu/workp/swp03011.pdf

Davis, P. M., Lewenstein, B. V., Simon, D. H., Booth, J. G., Connolly, M. J. L. (2008) "Open access publishing, article downloads, and citations: randomised controlled trial", British Medical Journal, 2008. Electronic copy available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2492576/

Dewatripont, M., Ginsburg, V., Legros, P., Walckiers, A., Devroey, J.P., Dujardin, M., Vandooren, F., Dubois, P., Foncel, J., Ivaldi, M, Heusse, M.D. (2006) "Study on the economic and technical evolution of the scientific publication markets in Europe", European

Commission. Electronic copy available at: http://ec.europa.eu/research/science-society/pdf/scientific-publication-study_en.pdf

Dewatripont, M., Ginsburgh, V., Legros, P., Walckiers, A. (2007) "Pricing of scientific journal and market power", Journal of the European Economic Association, Volume 5, Issue 2-3, 400–410, April-May 2007. Electronic copy available at: http://onlinelibrary.wiley.com/doi/10.1162/jeea.2007.5.2-3.400/abstract.

Dubini, P., Giglia, E. (2009) "Economic Sustainability during transition: the case of scholarly publishing", ELPUB Conference, Milano, June 2009.

Duranceau, E.F. (1995) "The economics of electronic publishing", Serial Review, pp-77-90.

European Commission Communication 'A Digital Agenda for Europe' (COM(2010)245, 19.05.2010) http://ec.europa.eu/information_society/digital-agenda/documents/digital-agenda-communication-en.pdf

European Commission Communication 'Innovation Union' (COM(2010)546, 06.10.2010) http://ec.europa.eu/research/innovation-union/pdf/innovation-union-communication_en.pdf

European Union Riding the Wave: How Europe Can Gain From The Rising Tide of Scientific Data, EU working group, 6 Oct. 2010 http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf

Edgar, B. D., Willinsky, J. (2010) "A survey of the scholarly journals using Open Journal Systems", Scholarly and Research Communication (1) 2. Available at http://journals.sfu.ca/src/index.php/src/article/viewFile/24/40

Edlin A. S., Rubinfeld, D. L. (2004) "Exclusion or Efficient Pricing: The "Big Deal" Bundling of Academic Journals", Antitrust Law Journal. Electronic copy available at: http://works.bepress.com/aaron_edlin/37

Eisenmann, T., Parker G., van Alstyne, M.W. (2006) "Strategies for two sided markets", Harvard Business Review, vol. 84, issue 10, 2006.

Fisher, J.H. (1997) "Comparing electronic journals to print journals: Are there cost savings?", Scholarly communication and technology. Conference organized by The Andrew W. Mellon Foundation at Emory University, April 24, 1997.

Fisher, J.H. (2008) "Scholarly Publishing Re-invented: Real Costs and Real Freedoms", The Journal of Electronic Publishing, Vol. 11, Issue 2, Spring 2008. Electronic copy available at: http://quod.lib.umich.edu/cgi/t/text/text-idx?c=jep;view=text;rgn=main;idno=3336451.0011.204

Fox, P. (2002) "Archiving of electronic publications – some thoughts on cost", Guest editorial, Learned Publishing, Vol. 15, no 1, 2002.

Frazier, K. (2001) “The Librarians' Dilemma: Contemplating the Costs of the "Big Deal", D-Lib Magazine, 7 (3). Electronic copy available at http://dx.doi.org/10.1045/march2001-frazier.

Gargouri Y., Hajjem C., Larivière V., Gingras Y., Carr L., Brody T, Harnad S. (2010) “Self-Selected or Mandated, Open Access Increases Citation Impact for Higher Quality Research”, PLoS ONE 5(10): e13636. doi:10.1371/journal.pone.0013636

Gentil Beccot, A.-Mele, S. Brooks T 2009, “Citing and Reading Behaviours in High-Energy Physics. How a Community Stopped Worrying about Journals and Learned to Love Repositories, arXiv”, http://arxiv.org/abs/0906.5418

Getz, M. (1992) “Electronic Publishing: An Economic View” Serials Review. Vol. 18, no 1-2, p. 25-31.

Giglia, E. (2010) “The Impact Factor of Open Access journals”, ELPUB Conference Milano. Electronic copy available at http://elpub.scix.net/cgi-bin/works/Show?_id=102_elpub2010&sort=DEFAULT&search=%2fseries%3a%22ELPUB%3a2010%22&hits=36

Ginsparg, P. (1997) “First Steps towards Electronic Research Communication”, Gate-ways to Knowledge: The Role of Academic Libraries in Teaching, Learning, and Re-search by Lawrence Dowler. Cambridge & London: MIT Press, p. 43-58.

Halliday L., Oppenheim, C. (2000) “Comparison and evaluation of some economic models of digital-only journals”, Journal of Documentation, Vol. 56, Issue 6, p. 660-673.

Halliday, L., Oppenheim, C. (2000) “Economic models of digital only journals”, Serials, Vol. 13, no 2, p. 59-65.

Harnad, S., Brody T., Vallières. F., Carr, L., Hitchcock S. et al. (2004) “The Access/ Impact Problem and the Green and Gold Roads to Open Access”, Serials Review, Vol. 30, p. 310–314. Electronic copy available at: http://www.sciencedirect.com/science/article/pii/S0098791307001530

Hawkins B.L. (1998) ”The unsustainability of the traditional library and the threat to higher education” in Hawkins, B. L., Battin, P., (eds) The mirage of continuity: reconfiguring academic information resources for the 21st century, Washington, DC: Council on Library and Information Resources, 1998. Electronic copy available at: http://lib.colostate.edu/images/about/goals/it/LibraryUnsustainability.pdf

Hedlund T., Gustafsson T., Björk B.C. (2004) ”The open access scientific journal: an empirical study”, Learned Publishing, Volume 17, Number 3, 1 July 2004, 199-209. Electronic copy available at https://helda.helsinki.fi/bitstream/handle/10227/603/hedlund_gustafsson_bjork.pdf?sequence=3.

Houghton, J. W. (2001) ”Crisis and transition: the economics of scholarly communication”, Learned Publishing, Volume 14, Number 3, July 2001, 167-176. electronic copy available at: http://www.ingentaconnect.com/content/alpsp/lp/2001/00000014/00000003/art00002.

Houghton, J.W., (2009a) “Open Access: What are the Economic Benefits? A Comparison of the United Kingdom, Netherlands and Denmark”, June 22, 2009. Electronic copy available at SSRN: http://ssrn.com/abstract=1492578

Houghton, J.W., Rasmussen, B., Sheehan, P.J., Oppenheim, C., Morris, A., et al. (2009) “Economic Implications of Alternative Scholarly Publishing Models: Exploring the Costs and Benefits”, London & Bristol: The Joint Information Systems Committee (JISC).

Houghton, J.W., Steele C., Sheehan, P.J. (2006) “Research Communication Costs in Australia, Emerging Opportunities and Benefits”, Department of Education, Science and Training, Canberra. Electronic copy available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1003951.

Houghton, J.W., Rasmussen, B. and Sheehan, P.J. (2010) Economic and Social Returns on Investment in Open Archiving Publicly Funded Research Outputs, Report to SPARC by Victoria University's Centre for Strategic Economic Studies. Electronic copy available at http://www.cfses.com/FRPAA/

Imboden, D.M. (2009) “Scientific Publishing: the dilemma of research funding organizations”, European Review, Vol. 17, No. 1, 23–31. Electronic copy available at: http://journals.cambridge.org/action/displayFulltext?type=1&fid=4324632&jid=ERW&volumeId=17&issueId=01&aid=4324624

Jeon, D.S., Menicucci, D. (2006) “Bundling electronic journals and competition among publishers”, Journal of the European Economic Association, Volume 4, Issue 5, 1038–1083, September 2006. Electronic copy available at: http://onlinelibrary.wiley.com/doi/10.1162/JEEA.2006.4.5.1038/abstract.

Jeon, D.S., Rochet, J.C. (2007) “The pricing of academic journals: a two-sided market perspective” Economics working paper 1025 Universitat Pompeu Fabra . Electronic copy available at: http://idei.fr/doc/by/rochet/pricingacademic.pdf

Katz M.L., Shapiro, C. (1985) “Network Externalities, Competition, and Compatibility”, The American Economic Review, Vol. 75, No. 3 (Jun., 1985), pp. 424-440. Electronic copy available at http://dep.eco.uniroma1.it/~miceli/Corsi-web/Miceli_EOI_Web/EOI-References/Katz-ShapiroAER85.pdf

Kaufman-Wills, “The facts about Open Access, ALPSP 2005 available at: http://www.alpsp.org/ngen_public/article.asp?id=200&did=47&aid=270&st=&oaid=-1

King, D.W. (2007) “The costs of journal publishing: a literature review and a commentary” Learned Publishing, Volume 20, no 2, April 2007, p. 85-106.

King, D. W., & Alvarado-Albertorio, F. M. (2008) "Pricing and other means of charging for scholarly journals: A literature review and commentary" Learned Publishing, Volume 21, Number 4, October 2008, pp. 248-272. Electronic copy available at: http://www.ingentaconnect.com/content/alpsp/lp/2008/00000021/00000004/art00002

King, D.W., Tenopir C. (1998) "Economic Cost Models of Scientific Scholarly Journals", Proceedings of ICSU Workshop, University of Oxford, 31 March - 2 April.

Kling, R., McKim, G. (2000) "Not Just a Matter of Time: Field Differences and the Shaping of Electronic Media in Supporting Scientific Communication", Journal of the American Society for Information Science, Vol. 51, 2000, p. 1306–1320.

Laakso M, et al. (2011) The Development of Open Access Journal Publishing from1993 to 2009. PLoS ONE 6(6): e20961. doi:10.1371/journal.pone.0020961

Lawrence, S. (2001) "Free online availability substantially increases a paper's impact", Nature, Vol. 411, Issue 6837, 2001, p. 521. Electronic copy available at: http://ivyspring.com/steveLawrence/SteveLawrence.htm

Lewis, D.W. (2004) "A demand-side view of the future of library collections", CNI, Fall 2004, Task Force Meeting. Electronic copy available at: http://www.cni.org/tfms/2004b.fall/

Marks, R.H., Duranceau, E.F. (1995) "The economic challenges of publishing electronic journals", Serial Review, Spring 95, Vol. 21, Issue 1, 1995.

Mabe, M., Amin, M. (2001) "Growth dynamics of scholarly and scientific journals", Scientometrics Volume 51, Number 1, 147-162. Electronic copy available at: http://www.springerlink.com/content/x235508307528q91/

McCabe, M.J. (2002) "Journal pricing and mergers: a portfolio approach", The American Economic Review March 259:269.

McCabe, M.J., Snyder, C. (2005) "Open Access and Academic Journal Quality", The American Economic Review, Volume 95, Number 2, 453-458. Electronic copy available at: http://www.jstor.org/pss/4132864

Montgomery, C.H., King, K.W. (2002) "Comparing Library and User Related Costs of Print and Electronic Journal Collections. A First Step Towards a Comprehensive Analysis", D-Lib Magazine, Volume 8 Number 10, October 2002. Electronic copy available at http://www.dlib.org/dlib/october02/montgomery/10montgomery.html

Morris, S. (2005) "The true costs of scholarly journal publishing", Learned Publishing, Vol. 18, no 2, 2005, p. 115-126.

Odlyzko, A.M. (1997) "The economics of electronic journals", First Monday, Vol. 2, no 8, 1997.

Odlyzko, A.M. (1998) “The economics of electronic publishing”, Journal of Electronic Publishing, Vol. 4, Issue 1, 1998.

Patterson, M. (2011) “Open Access publishers: breaking even and growing fast”, APE Berlin 2011.

Peters, P.E. (1999) "Cost Centers and Measures in the Networked Information Value-Chain", Journal of Library Administration, Vol. 26, no 1/2, 1998, p. 203-212.

Prosser, D. C. (2005) “The Next Information Revolution - How Open Access will Transform Scholarly Communications”, electronic copy available at: http://eprints.rclis.org/bitstream/10760/6352/1/IYLIM_chapter.pdf;

Proudman, V. (2008) “The population of repositories”, Eds. K. Weenink, L. Waaijers and K. van Godtsenhoven, A DRIVER's Guide to European Repositories, p. 49 – 101. Amsterdam: Amsterdam University Press. Electronic copy available at: http://dare.uva.nl/aup/nl/record/260224

Read, J. M. (2003) “Cataloguing without tears: managing information in the information society” Oxford, Chandos Publishing.

Rochet, J.-C., Tirole J. (2003) “Platform Competition in Two-Sided Markets”, Journal of the European Economic Association, Vol. 1, 2003, p. 990-1029.

Rochet, J.-C., Tirole J. (2004) “Two-sided markets: an overview” Toulouse: Mimeo, IDEI.

Romary, L., Armbruster, C. (2010) “Beyond Institutional Repositories” International Journal of Digital Library Systems, Vol. 1, Issue 1, 2010, p. 44-61; Electronic copy available at: http://ssrn.com/abstract=1425692

Roosendaal, H.E., Geurts P. A. (1997) “Forces and Functions in Scientific Communication: An Analysis of their Interplay,” Conference on "Co-operative Research in Information Systems in Physics" September 1-3, University of Oldenburg, Germany. Electronic copy available at: http://doc.utwente.nl/60395/1/Roosendaal97forces.pdf

Rowland, F. (2002) “The peer-review process”, Learned Publishing, Vol. 15, no 4, 2002, p. 247-258.

Seringhaus M., Gerstein M. (2006) “The Death of the Scientific Paper”, The Scientist 2006, 20:25.

Seringhaus M., Gerstein M. (2007) “Publishing perishing? Towards tomorrow’s information architecture” BMC Bioinformatics, 8:17 doi:10.1186/1471-2105-8-17 Electronic copy available at: http://www.biomedcentral.com/1471-2105/8/17

Shieber, S. (2009) “What percentage of open-access journals charge publication fees?”, The Occasional Pamphlet, 2009. Electronic copy available at:

http://blogs.law.harvard.edu/pamphlet/2009/05/29/what-percentage-of-open-access-journals-charge-publication-fees/

Schonfeld, R.C., King D.W., Okerson A., Fenton G.E. (2004) "Comparison of non-subscription costs of print and electronic formats on a life-cycle basis", D-Lib Magazine, Vol. 10, January 2004.

SOAP 2011 "the Open Access publishing landscape" SOAP Symposium, Berlin January 13,2011 http://project-soap.eu/report-from-the-soap-symposium/

Suber, P. (2008) "Thinking about prestige, quality and open access", SOAN newsletter, September 2008.

Suber, P. (2009) "A field guide to misunderstandings about open access", SOAN newsletter April 2009 http://www.earlham.edu/~peters/fos/newsletter/04-02-09.htm

Suber P. (2009 a) "Ten challenges for Open Access journals" SOAN newsletter October 2009 http://www.earlham.edu/~peters/fos/newsletter/10-02-09.htm

Swan, A. (2008) "The business of digital repositories", Weenink K., Leo Waaijers L. van Godtsenhoven K., A driver's guide to European Repositories. Amsterdam: Amsterdam University Press.

Swan, A. (2010) "The Open Access citation advantage: Studies and results to date", Technical Report, School of Electronics & Computer Science, University of Southampton. Electronic copy available at: http://eprints.ecs.soton.ac.uk/18516/

Teece D.J. (2010) "Business models, business strategy and innovation" Long Range Planning 43, 2-3 April June 2010 172-194.

Tenopir, C., King, D.W. (2000) "Towards Electronic Journals: Realities for Scientists, Librarians and Publishers", Washington DC: Special Libraries Association.

Tenopir, C., King, D. W., Edwards, S., Wu, L. (2009) "Electronic journals and changes in scholarly article seeking and reading patterns", Aslib Proceedings: New Information Perspectives, 61(1), 5-32.

Varian, H. (1996a) "Differential pricing and efficiency", First Monday, Volume 1, Number 2 - 5 August 1996. Electronic copy available at: http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/473/829

Varian, H. (1996b) "Pricing Electronic Journals" D-Lib Magazine, June 1996. Electronic copy available at: http://webdoc.sub.gwdg.de/edoc/aw/d-lib/dlib/june96/06varian.html

Waltham, M. (2006) "Learned society business models and open access: overview of a recent JISC funded study", Learned Publishing, Vol. 19, no 1, January 2006, p. 15-30.

Way, D. (2010) “The Open Access Availability of Library and Information Science Literature”, College & Research Libraries, Vol. 71, 2010, p. 302-309. Electronic copy available at: http://works.bepress.com/doug_way/2/

Willinsky, J. (2003) “Scholarly Associations and the Economic Viability of Open Access Publishing”, Journal of Digital Information, Vol. 4, Issue 2. Electronic copy available at: http://www.biomedcentral.com/openaccess/archive/?page=features&issue=16

Willinsky, J. (2009) “The Stratified Economics of Open Access”, Economic Analysis and Policy 39,1, March, 53-70.

Other references

Antelman, K. (2004) “Do Open Access articles have greater research impact?”, College and Research libraries, Vol. 65, no 5, 2004, p. 372-382. Available at: http://eprints.rclis.org/archive/00002309/01/do_open_access_CRL.pdf

Benkler, Y. (2006) “The Wealth of Networks: How Social Production Transforms Markets and Freedom”, New Haven: Yale University Press.

Bide, M., Oppenheim C., Ramsden, A. (1997) “Charging Mechanisms for Digitised Text”, Second Supporting Study for the JISC/PA working party. Electronic copy available at: http://www.ukoln.ac.uk/services/elib/papers/pa/charging/

Blixrud, J., Jewell. T.D. (1998) "Understanding Electronic Resources and Library Materials Expenditures: An Incomplete Picture" ARL: A Bimonthly Newsletter of Research Library Issues and Actions. Vol. 197, 1998, p. 12-13.

Bird, C. (2008) “Oxford Journals’ adventures in open access” Learned Publishing, Vol. 21, no 3, 2008, p. 200-208.

CEPA. (2008). “Activities, costs and funding flows in the scholarly communications system in the UK” A report commissioned by the Research Information Network, London.

Courant P., J. O’Donnell, A. Okerson and C. Taylor. 2010. “Improving Access to Research”. Editorial. Science, Vol. 22, no 393.

Eysenbach, G. (2006) “Citation Advantage of Open Access Articles”, PLoS Biology, 4 (5), e157. doi:10.1371/journal.pbio.0040157. Electronic copy available at: http://www.plosbiology.org/article/info%3Adoi%2F10.1371%2Fjournal.pbio.0040157.

Gaule, P., Maystre, N. (2008) “Getting cited: does open access help?” JEL classification: O33, O38. CEMI-WORKINGPAPER-2008-007. Electronic copy available at: http://ssrn.com/abstract=1427763

Gerhard, K.H. (2005) "Pricing models for electronic journals and other electronic academic materials", Journal of Library Administration, Volume 42, Issue 3-4, 2005. Electronic copy available at: http://www.tandfonline.com/doi/abs/10.1300/J111v42n03_01.

Giglia, E. (2010) "Open access to scientific research: where are we and where are we going? Facts and figures on the occasion of the 2010 Open Access Week", European Journal of Physical and Rehabilitation Medicine, Vol. 46, no 3, p. 461-469.

Guédon, J-C. (2001) "In Oldenburg's Long Shadow: Librarians, Research Scientists, Publishers, and the Control of Scientific Publishing", Proceedings of the 138th ARL (American Research Libraries) Membership Meeting. Toronto, May 23–25. Electronic copy available at:
http://www.arl.org/resources/pubs/mmproceedings/138guedon.shtml

Hawley, J.B. (2004) "JCI & Open Access" Proceedings of the Conference on Open Access Publishing, Society for Scholarly Publishing, November 2004. Electronic copy available at:
http://www.sspnet.org/i4a/pages/index.cfm?pageid=3670

Harnad S., Brody T. (2004) "Comparing the Impact of Open Access (OA) vs. Non-OA Articles in the Same Journals", D-Lib Magazine, Vol. 10, no 6, 2004. Electronic copy available at: eprints.ecs.soton.ac.uk/10207/

Harnad, S. (1995) "A Subversive Proposal", Okerson A., O'Donnell J. (Eds.), "Scholarly Journals at the Crossroads; A Subversive Proposal for Electronic Publishing", Washington DC: Association of Research Libraries, June 1995. Electronic copy available at:
http://www.arl.org/sc/subversive/

Hendley, T. (1998) "Comparison of Methods and Costs of Digital Preservation", British Library Research and Innovation, Report 67. London: The British Library. Also available online at:
http://www.ukoln.ac.uk/services/elib/papers/tavistock/hendley/hendley.html.

Hesmondhalgh, D. (2007) "The Cultural Industries", 2nd ed. London: Sage.

Houghton, J.W., Sheehan P.J. (2009) "Estimating the Potential Impacts of Open Access to Research Findings" Economic Analysis and Policy. Vol. 39, no 1, 2009.

King, D.W., Griffiths, J-M. (1995) "Economic Issues Concerning Electronic Publishing and Distribution of Scholarly Articles", Library trends, Vol. 43, no. 4, pp 713-40,1995.

King, D.W., Xu, H. (2003) "Library consortia & electronic journal services", Pearce, A., "The Consortium license. Is it a sustainable model?" Oxford: Ingenta Institute.

King, D.W. (2004) "Should Commercial Publishers Be Included in The Model for Open Access through Author Payment?" D-Lib Magazine, Vol. 10, no 6, 2004.

Maron, N.L., Kirby Smith, K. (2008) "Current models of scholarly communications"

Washington: ARL, 2008. Electronic copy available at:
http://www.arl.org/bm~doc/current-models-report.pdf

Maron, N.L. (2009) "Hindawi publishing Corporation: the open access contributor pays model" JISC, as part of the Ithaka Case Studies in Sustainability project. Electronic copy available at:
http://www.ithaka.org/ithaka-s-r/research/ithaka-case-studies-in-sustainability/case-studies/SCA_BMS_CaseStudy_Hindawi.pdf

McVeigh, ME. (2004) "Open Access Journals in the ISI Citation Databases: Analysis of Impact Factors and Citation Patterns." A Citation study from Thomson Scientific. Electronic copy available at:
http://science.thomsonreuters.com/m/pdfs/openaccesscitations2.pdf

King, C.J.; et al. (2006) "Scholarly communication: academic values and sustainable models", UC Berkeley, CSHE Papers, 2006. Electronic copy available at:
http://escholarship.org/uc/item/4j89c3f7

Kirsop, B. (2007) "Before and After: OA in Developing Countries" Presentation delivered at Berlin 5 Open Access: from Practice to Impact: Consequences of Knowledge Dissemination, Padova.

Nelson, R.R. (2004) "The market economy, and the scientific commons", Research Policy, Vol. 33, 2004, p. 455-471. Electronic copy available at: http://ideas.repec.org/p/ssa/lemwps/2003-24.html

Norris M., Oppenheim, C., Rowland, F. (2008) "The Citation Advantage of Open-Access Articles", Journal of the American Society for Information Science and Technology, Vol. 59, 2004, p. 1963–1972.

PRISM, Partnership for Research Integrity in Science and Medicine, 2007. Available at
http://www.prismcoalition.org/.

Repositories Support Project 2010. "RSP - Repository Software Survey, March 2009."

RIN 2011. Heading for the open road: costs and benefits of transitions in scholarly communications. Electronic copy available at: http://www.rin.ac.uk/our-work/communicating-and-disseminating-research/heading-open-road-costs-and-benefits-transitions-s

Saloner, G., Shepard, A., Podolny, J. (2001) "Strategic management", New York: Wiley

Shapiro, C., Varian, H. (1999) "Information Rules. A Strategic Guide to the Network Economy", Boston: HBS Press.

Schonfeld, R.C., King, D.W., Okerson A., Fenton, G.E. (2004a) "The non-subscription side of periodicals: changes in library operations and costs between print and electronic journals" Washington DC: Council on Library and Information Resources.

Suber, P. (2002) "Open Access to the Scientific Journal Literature", Journal of Biology, Vol. 1, Issue 1, June 2002.

Suber, P. (2004) "What is Open Access? An Overview" Proc. Conf. Open Access Publishing, Society for Scholarly Publishing, November 2004, slide-set at: http://sspnet.org/home.aspx?404;http://www.sspnet.org:80/i4a/pages/index.cfm?pageid=3674

Suber, P. (2005) "Open access, impact, and demand. Why some authors self archive their articles", BMJ. 2005 May 14; 330(7500): 1097–1098. Electronic copy available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC557876/pdf/bmj33001097.pdf

Suber, P. (2007). "Flipping a journal to open access" SPARC Open Access Newsletter 114 (October 2). Electronic copy available at: http://www.earlham.edu/~peters/fos/newsletter/10-02-07.htm#flip.

Sun, M., Tse, E. (2007) "Sustainable growth of payment card networks: A two sided market approach", Journal of Business Strategies, Vol. 24, 2007, p. 165-191.

Universities UK and R.I.N. (2009) "Paying for open access publication charges", available at www.rin.ac.uk/openaccess-payment-fees.

Weitzman, J.B., Lessig L. (2004) "Open Access and creative common sense", "Open Access now", 14 May 2004. Electronic copy available at: http://www.biomedcentral.com/openaccess/archive/?page=features&issue=16

Willinsky J., Mendis R. (2007) "Open access on a zero budget: a case study of postcolonial text", Information Research, Vol. 12, Issue 3, paper 308. Electronic copy available at: http://www.aughty.org/pdf/postcolonial_text.pdf

Young, N.S. et al. (2008) "Why Current Publication Practices May Distort Science" PLoS Medicine, Vol. 5, Issue 10, 2008, e201. 5 (10), 2008, e201. Electronic copy available at: http://www.plosmedicine.org/article/info:doi/10.1371/journal.pmed.0050201

All links are active as of September 20, 2011.

Annex 1: Publishers characteristics

Table 1: Companies description

Data	Total journals	Journals with OA option	Total journals owned	Total employees	Printed books in catalogue	Ebook titles	Business areas	Disciplines in journals	Geographic scope
BMJ	32	1	100%	383	4	n.a.	Journals, Evidence-based medicine tools, learning material	Medicine and Healthcare	England, USA, Asia, Europe
CAMBRIDGE	239	20	48%	n.a.	36.000	8.898	Books, journals, learning materials, bibles	Humanities, Science, Social sciences, Technology, Medicine	Europe, Middle East and Africa (EMEA), Americas, Asia-Pacific
EDP SCIENCES	47	9	30%	60	550	180	Journals	STM, Social sciences, Humanities, General knowledge	France and Morocco
ELSEVIER	2.392	72	75%	7.000	19.603	6.297	Books, journals, events	Science, Health information	24 countries
IOP	67	6	43%	302	n.a.	n.a.	Journals	Physics	UK, USA, China, Russia, Japan
NATURE	68	25	41%	800	n.a.	n.a.	Journals, events	Chemistry, Clinical practice and research, Earth and environment, Life science, Physics	Main offices in London, New York and Tokyo
SAGE	560	Open Access Option	62%	13.600	700	1.400	Books, journals, learning resources	Business, Humanities, Social sciences, Science, Technology, Medicine	Los Angeles, London, New Delhi, Singapore, Washington DC
SPRINGER	2.000	17 with processing fee	65%	5.000	45.322	37.000	Books, journals	Science, Technology, Medicine, Business, Transport, Architecture	20 countries
TAYLOR AND FRANCIS	1.500	320	n.a.	8.000	20.000	16.000	Publishing, events, commercial and services	Behavioural science, Humanities and Social sciences, Science	20 global offices
WILEY-BLACKWELL	1.500	588	50%	1.725	25.000	9.000	Journals, books, professional learning	Agriculture, Medicine, Business and Management, Social sciences, Mathematics, STM, Health, Physics, Education	US, Canada, Germany, Denmark, Australia, Delhi, Singapore, Tokyo, Beijing, Shanghai

Table 2 Information gathered on individual publishers

Data	Notes	BMJ	CAMBRIDGE	EDP SCIENCES	ELSEVIER **	IOP	NATURE	PLOS	SAGE	SPRINGER	TAYLOR AND FRANCIS	WILEY-BLACKWELL
Nature of publisher	institutional form (company, not for profit)	X	X	X	X	X	X	X	X	X	X	X
Total journals		X	X	X	X	X	X	X	X	X	X	X
n. of OA journals		X	X	X	X	X	X	X	X	X	X	X
total employees		X		X	X	X		x		X	X	X
total books in catalogue		X	x	x	X	x	x		x	x	X	X
ebook titles			x	x	X	x	x			x	X	X
new ebook titles						x	x			x		X
business areas		X	x	x	X	x	x	x	x	x	X	X
disciplines in journals		X	x	x	X	x	x	x	x	x	X	X
countries	countries where the company has headquarter, sales companies, divisions, owned or controlled companies	X	x	x	X	x	x	x	x	x	X	X
externalised activities	activities oursourced to third parties that are not owned or controlled by the companies	X		x	X	x	x			x		X
Where	countries where outsourced activities are located			x	X	x				x		X
%Total journals owned	% of journals owned and published by the companies	X	x	x	X	x	x	x	x	x		X
of which peer review organized in house	% of owned journals for which peer review is organized by the company	X	x	x	X	x	x	x		x		X
Tracking system	existence of a tracking system visible to the author	X	x	x	X	x	x	x	x	x	X	X
n. of people involved in peer review organization	FTE (ful time equivalent) employed people involved in the management of the peer review process (i.e. employed editors)	X		x	X	x	x	x	x	x		X
salaries for peer review organization	total company cost for salaries to FTE involved in peer review organization	X		x		x		x	x			X
external cost	total costs paid to external suppliers (for instance editors not employed by the company) associated with organization and management of peer review	X		x		x			x			X
n. of articles	total number of artiche currently stored in digital repository	X	x	x	X	x		x	x	x	X	X
other documents	total number of other documents (for insstence ebooks) currently stored in digitl repository			x	X	x		x		x		X
n. of new articles per year	number of new articles added to repository each year	X	x	x	X	x		x	x	x		X
n. of dedicated people	number of FTE employed by the company to manage activities associated with article publication (typesetting, cross referencing, translation into HTML….)	X	x	x		x		x	x	x		X
total cost employed people	total company costs for salaries of FTE associated with article publication	X		x		x			x			X

total cost for externalized services	total costs paid to external suppliers (for instance third parties in India) in charge of article publication	X		x		x			x			X
last update of the platform	year of the last significant investment on the publication platform	x	x	x		x		x	x			X
total investment	total investment on the platform so far (if this datum is not available please indicate the amiunt of the last investment made on the platform			X	X	X		x		X		X
total salaries for platform/ IT management	total company salary costs for employees associated with platform management	X		x	X	x						X
maintenance costs	total yearly maintenance costs (excluding employed personnel) for platform	X		x	X	x						X
PEER compliance cost	total costs and investments necessary to comply with PEER requirements	X	x	x		x			x	x		X
n. of journals in PEER		X	x	x	X	x	x		x	x	X	X
sales forces	number of people in charge of sales (might be agents or employed by the company)	X	x	x		x			x			X
salaries for salesforce	total company salary costs for salesforce	x	x	x		x						X
total marketing and sales costs	total costs for marketing and sales EXCLUDING salaries for salesforce			x		x		x		x		
% of revenues from journals on tot. company revenues		x	x	x		x		x				X
% revenues books			x	x		x				x		X
% other revenues			x	x		x						X
consolidated revenues		x	x	x	X	x		x		x	X	X
operating profit					X	x		x		x	X	X

** Elsevier did not provide data as requested; however it gave us costs calculated internally with a full cost technique and was extensively available for comments and clarifications

Annex 2: Repositories characteristics

Table 1. Repositories characteristics

	Year of set up	Type of repository	Nature of institution	Number of research units/departments	Researchers involved	Disciplines
Cornell University (ArXiv)	1991	Disciplinary	University	n.a.	n.a.	Mathematics, Physics, Astronomy, Computer Science, Quantitative biology and statistics.
CSIC	2002	National	Governmental.	138	6.000	Humanities and social sciences, Biology and biomedicine, Natural resources, Agricultural sciences, Physical science and technologies, Materials science and technology, Food science and technology, Chemical science and technology
Gesis	2008	Disciplinary	Indipendent, Publicly prefunded	5	250	Social Sciences
GoeScholar	2009	Institutional	University	13	3.246	Philosophy and Theology, Medicine, Mathematics and informatics, Social sciences
HAL	2005	National	Governmental.	210	2.800	Applied Mathematics, Computation and Simulation; Algorithmic, Programming, Software and Architecture; Networks, Systems and Services, Distributed Computing; Perception, Cognition, Interaction; Computational Sciences for Biology, Medicine and the Environment.
Max Planck (eDoc)	2002	Institutional	Independent non-governmental and non-profit	80	5.222	Natural sciences, life sciences, social sciences, and the arts and humanities
NIH (PubMed Central)	1996	Institutional	National Institutes of Health	n.a.	n.a.	Health Sciences
SSRN	1994	Disciplinary	Private Website	n.a.	n.a.	Social sciences, Humanities
Università degli Studi di Milano (A.I.R. Institutional Archive for Research)	2005	Institutional	University	6	8.663	Pharmaceuticals, Law, Literature and Philosophy, Medicine, Natural, Mathematics and Physics

Table 2 Information gathered on individual repositories

		CSIC	Gesis	Göttingen University	Inria/CNRS	Max Planck Society	Università degli Studi di Milano
Type of repository		x	x	x	X	X	x
Date of set up		x	x	x	X	X	x
Number of researchers involved	researchers working in the institution (institutional); researchers addressed (national)	x	n.a.	x	X	X	x
Material archived							
# of references archived	total number of references	x	x	x	X	X	x
new references per year	references added to repository	x	x	x		x	x
format of references	pdf vs xml …	x	x	x	X	x	x
full text searchability		x	x	x		x	x
nature of reference	composition of reference by type of information						
% publication records		x	x	x	X	x	x
% publication record + link		x	x	x	X		x
% full text		x	x	x	X	x	x
total number of full text documents		x	x	x	X	x	x
type of material archived	composition of reference by stage of publication						
% unpublished material	includes preprint, PhD dissertations, teaching and research material	x	x	x	X	x	x
% published material	stage 2 or stage 3 articles	x	x	x	X	x	x
feeding the archive							
% of references self archived by authors		x	x	x	X	n.a.	x
% of references inserted by staff		x	x	x	X	x	x
% of references provided by publishers		x	x	x	X	x	x
% of references harvested from other respositories		x	x	x	X	x	x
FTE employed to collect material		x*	x	x	X	x	x
Cost for FTE	if not provided, can be estimated	x*	x	x	X		x
# of FTE involved in processing material	metadata, editing, data entry, quality control, embargo checking..	x*	x	x	X	n.a.	x
Cost for FTE		x*	x	x	X		x
Cost for externalized processing services	is there any? If yes amount		x	x	X	x	x
# of FTE employed in support activities	is there anybody in charge of supporting authors who are self archiving?	x*	x	x	X	x	x

Cost for FTE to support authors		x*	x	x	X	x	x
Initial cost for set up (hardware)	is the initial investment known or traceable?	x	x	x	X	x	x
External cost for software development	is there any? Within the institution or externalised? If known, amount	x	x		X	x	x
FTE involved in software development		x*	x	x	X	x	x
Cost for FTE Software development		x*	x		X		x
Yearly maintenance cost	cost for server and maintenance	x	x	x	X	x	x
Long term preservation	internal vs outsourced			x	X	x	x
FTE dedicated to PEER repository				x	X	x	
cost of FTE				x	X	x	
setup investment	one off costs			x	X		
maintenance costs				x	X		

* number of people and costs provided as a total.