UK scholarly journals: 2006 baseline report

An evidence-based analysis of data concerning scholarly journal publishing

Final report

Prepared on behalf of the Research Information Network, Research Councils UK and the Department of Trade & Industry

By Electronic Publishing Services Ltd
http://www.epsltd.com, EPS is an Outsell, Inc company

In association with Professor Charles Oppenheim and LISU at Loughborough University Department of Information Science (LUDIS)

12 September 2006

www.rin.ac.uk www.rcuk.ac.uk www.dti.gov.uk

Contents

Page

Foreword from RIN, RCUK and DTI

The UK science base is one of the most productive and influential in the world. The Government's ambition is to build on this success, to sustain the UK's strengths in high-quality research, and to enhance the impact of research in promoting innovation and improving the quality of life in the UK and overseas. Effective communication of the results of research is of fundamental importance in realising these ambitions. High-quality research demands high-quality scholarly communications.

The UK has a long tradition in scholarly publishing. The first scholarly journal in the world, the *Philosophical Transactions* of the Royal Society, was published in London in 1665. Since then, scholarly journals have come to play a central and critical role in scholarly communications, and the UK has sustained its position as one of the leaders in the global scholarly communications business.

The fundamentals of the ways in which that business operated, and the roles that the key groups of players performed, changed little for over three hundred years after 1665. But the internet has in recent years begun to shape revolutionary change, as a truly disruptive technology. New ways of communication between researchers have begun to challenge inherited assumptions about the roles of the key groups of players and stakeholders – researchers, research funders, publishers, librarians and others – about how they can most effectively perform those roles, and indeed about the key purposes they are seeking to fulfil.

As researchers, publishers and librarians have considered how best to respond to these challenges, the open access debate has generated heat as well as light. The debates have often been characterised by misunderstandings, and there have been tensions over the quality and completeness of the basic information that the different stakeholders have presented in support of their respective positions.

The key issue of public policy is how best to promote and support, in this rapidly-changing digital world, the continuing and sustainable development of a world-class scholarly communications system for the UK. In order to develop policies that will meet those goals, we need solid and reliable evidence about where we are now. It was to achieve at least part of that end that the Research Information Network (RIN), the Department of Trade and Industry (DTI) and Research Councils UK (RCUK) came together in late 2005 to commission a thorough review of data relating to the operation and costs of scholarly journal publishing, with an emphasis on data concerning the UK.

In order to ensure that the results of the study should be as authoritative as possible, the sponsors put together an expert panel of representatives of each of the main groups of stakeholders. We gave to them the collective and collaborative role of rigorously questioning and checking the data and the conclusions presented to them. This report therefore presents results that have been subject to rigorous scrutiny by representatives of the research, publishing and library communities, as well as by the sponsors. This is the first time that such an approach has been adopted in work of this kind, and we are confident that the results will be accepted as an authoritative account of the current state of knowledge about scholarly journal publishing in the UK.

The report presents, therefore, as authoritative a base of evidence as can currently be constructed in the key areas of:

- *the volume and value of the academic journal market*, including such key issues as the splits between commercial and learned society publishers,

and between print and electronic revenues; acquisition and cancellation trends; the split of publisher revenues between academic, corporate and personal subscribers; and the value of pay-per-view sales

- *journal supply-side economics,* including the effort and costs incurred by researchers and by publishers as part of their contribution to the publishing process; any differences between commercial and learned society publishers; and costs of launching new products
- *usage,* including the split between the leading journals and the rest; the extent and potential of unmet demand; and the barriers to that demand being met
- *citations and impact factors,* and such key issues as whether articles in subscription journals are more likely to be cited than those in open access journals, or vice versa; any resultant variation in impact factors; and any relationships between citations and impact factors on the one hand, and large-scale collaborative research on the other
- *disciplinary differences,* and whether there are significant differences between researchers as readers or authors in different disciplines
- *costs and impact of open access journals and of digital repositories,* and the differences between these new models and of publishing conventional journals.

In each of these areas there are issues and questions to which we should ideally like to have answers, but where we have very little data which can be described as authoritative. One of the important results of the study is to identify such gaps, and to point to ways in which they might be filled. Like many studies, therefore, the current one points to the need for further work. Nevertheless, we believe that it already constitutes an important contribution towards creating the kind of authoritative evidence base which is needed for effective policy and decision-making.

The RIN, RCUK and the DTI are very grateful to EPS Ltd. who undertook the study, and also to the members of the expert panel – chaired by Professor Jeffrey Aronson of the University of Oxford – who gave so freely of their time and expertise in producing this report. We shall be using the results of this work in working with all the key groups of stakeholders in our continuing work to try to ensure that the UK does indeed continue to support its world-class researchers with the best possible scholarly communications system.

Preface from the Chair of the Expert Panel

"Publish and be damned" was what the Duke of Wellington supposedly scrawled across a letter that he received from a publisher called John Joseph Stockdale, who threatened to reveal details about the Duke's relationship with a high-class courtesan called Harriette Wilson (née Dubouchet), also known as Mrs Q. The Duke was above such threats. Stockdale began publishing Harriette's memoirs in 1825, to huge demand. Although her account was name-dropping rather than salacious, her sensational opening sentence is justifiably famous: "I shall not say why, or how, I became, at the age of fifteen, the mistress of the Earl of Craven."

"Publish and be sued" was the spin that Richard Ingrams later put on the original, reflecting the hazards of being Editor of the British satirical magazine *Private Eye*. "Publish and be charged" is the modern version. The traditional publishing model for research, especially in the biosciences, has been publication in scholarly journals without [usually] a charge on the researcher, the costs being borne by the reader/subscriber. However, the author-pays model of "open-access" publishing threatens to overturn that tradition. And this is just one of many issues that the advent of the internet has imposed on scholarly publishing.

Now, as a clinician faced with a sick patient I consider the evidence before embarking on a series of investigations or a course of treatment. My patient would hardly thank me for proceeding in the dark. But when it comes to public policy, the policy makers, even when they are trained scientists, often seem to be prepared to proceed without any such review, or indeed without any evidence, simply relying on opinions. I was therefore pleased to be invited to chair a panel of experts who had been asked to advise the authors of this report as they gathered the evidence relating to important aspects of modern journal publishing. The Expert Panel, and other experts who were consulted, came from the library, research funder, publisher, and research communities; as would be expected, they generated lively, constructive debate about the interpretation of key data.

It was the ambition of those undertaking this project to produce a practical, evidence-based report on key issues in journal publishing, in order to clear the ground for debate, further research, and consequent policy development. In order to achieve this, their approach had to be rigorous, objective, and scholarly. This was a challenging brief, but they have met it. Their comprehensive filtering and critical evaluation of the relevant data sources distinguishes this report from others that have preceded it. The Expert Panel was particularly struck by the paucity of hard evidence in this important field and in particular the scarcity of randomized studies. Never has that trite phrase "more research is required" been more readily applicable.

We hope that this report, which we believe is a thorough and balanced view of the current evidence base, will guide future debate and research in this area. We are sure that the output will be a useful tool for all stakeholders focussed on improving scholarly communications in the UK.

And to the authors of the report we say, "Publish and be congratulated."

Professor Jeffrey Aronson, British Pharmacological Society.

About the report authors

This report has been prepared by Electronic Publishing Services Ltd (EPS) in association with Professor Charles Oppenheim and LISU, both located at Loughborough University Department of Information Science (LUDIS).

Individual authors for key report areas are as follows:

1. Journal market value and volume – Leonie Nicholas, Consultant, EPS;
2. Journal supply-side economics – Leonie Nicholas, Consultant, EPS;
3. Usage – Eric Davis and Suzanne Lockyer, LISU;
4. Citations, impact factors and their role – Professor Charles Oppenheim, LUDIS;
5. Disciplinary differences– Eric Davis and Suzanne Lockyer, LISU;
6. Cost and impact of alternative formal dissemination models – David Powell, Consultant, EPS.

Declaration of conflict of interest

Since being commissioned to produce this report, Electronic Publishing Services (EPS) Ltd. has been acquired by US consultancy firm Outsell, Inc. Both EPS Ltd. and Outsell, Inc. are authors of key studies evaluated and cited in this study. This conflict of interest was declared at the outset of the project and subsequent to EPS' change in ownership. Objectivity has been sought throughout via the external Expert Panel assigned to the project who have reviewed deliverables at key points in the report preparation process.

Executive summary

Introduction

The objective of this report is to audit and evaluate the available evidence relating to scholarly journals publishing, and furthermore to distil what we do and do not currently know for certain about this market.

The report provides a snapshot of the evidence base about the scholarly journal publishing process and seeks to provide a basis of shared knowledge which policy-makers and research funders can draw on to consider if and how the system might develop, to ensure that the research publishing process delivers best value and best practice.

Area 1: Journal market volume and value

Key questions

The particular issues identified for review in this section are:

1. Industry volume and value
2. Split between commercial and society publishers, volume and value
3. Acquisition and cancellation trends
4. Split between print and electronic revenues
5. Split of publisher revenue between corporate, academic and personal subscribers
6. Value of PPV article sales

The evidence

Volume

There are estimated to be around 20,000-25,000 peer-reviewed scholarly journals actively being published worldwide. The volume of journals has grown consistently at a compound annual growth rate of 3-4% over the past one hundred years.

An estimated 60% of all journals are published online (most are also parallel published in print). This figure is reported to be as high as 90% for English-language journals emanating from US/UK-located publishers. Around 10% of peer-reviewed scholarly journals are published under some form of open access model.

There are estimated to be at least 200,000 full time researchers working across industry, government and higher education in the UK. Global sources suggest that there may be 5.5 million researchers globally.

Value

Serial subscriptions spend by UK higher education institutions (of which peer-reviewed scholarly journals are a sub-set of unknown size) totalled £94.5 million in 2003-2004. The number of current serial subscriptions has doubled in higher education institutions over the past ten years. Much of this growth in volume is thought to be attributable to 'bundled' or packaged deals offered by publishers.

Published sources estimate publisher revenues of circa $5 billion for English-language journals in scientific, medical and technical disciplines in 2004. It is likely that this is a significant under-estimate of the total peer-reviewed journal market (it excludes non-English language journals, journals in disciplines outside of STM such as social sciences, the arts and humanities and is necessarily based on 'best estimates' of journal revenues derived from figures in published accounts).

The customer profile will vary from journal to journal depending on business model and its disciplinary focus. Overall it is likely that academic customers account for the largest proportion of these revenues (c30-40% of revenues), with significant numbers of corporate customers (c20-30% of revenues), and government, individual and professional subscribers (e.g. healthcare professionals) and advertising clients (such as pharmaceutical companies) accounting for the balance.

The gaps

The somewhat dislocated story above reveals much about the gaps that undermine the existing baseline data about scholarly journals.

In terms of the market constituents, we have particularly poor data about buyers of scholarly journals beyond the higher education customers. Even basic data such as the numbers of these types of special libraries and information centres which exist in the UK, are not published.

Further work on sizing and segmenting the publisher market would also be valuable – we have indicated how sensible estimates can potentially be derived from available sources such as trade associations.

The general nature of published reports regarding the size and structure of the journals publishing market hinders our interpretation of important trends in this marketplace. Studies gathering data relevant to the STM publishing market are too broad in product focus and too narrow in terms of disciplinary coverage to deliver in-depth data relevant to the market defined as the subject of this report.

Deeper analyses of the value of scholarly journal publishing (such as sales to different types of customers, sales in different formats and the balance of different revenue streams) are not possible based on the available evidence.

More survey data directly sourced from the publishers themselves (anonymised and aggregated) and tracked over time would go a long way to filling the gaps in the evidence base on market size and structure.

Area 2: Journal supply-side economics

Key questions

The key issues under examination in this area of the report are:

1. Evidence about the effort and cost incurred on the one hand by publishers and on the other by academics as part of their respective contributions to the publications process (for instance, costs associated with editorial and quality issues – including both the external and internal aspects of peer review).

2. Data that might help in developing understanding of any differences that there may be in journal economics between commercial and learned society publishers.
3. Reliable data on the costs of launching new products (both direct and indirect) and on the factors determining the investment required for new launches.

The evidence

Estimates suggest that 'content creation' costs, sometimes called 'first copy costs' (which do not vary with either the volume of output or the number of subscribers served) vary widely from publisher to publisher with data points collected ranging from $250 to $2,000 per published article. Baseline data relating to specific parts of the process are rare, including specific costs relating to peer review.

In addition to these 'first copy article costs', publishers incur both variable costs relating to production and distribution and general fixed overheads. Variable costs for journal publishing are very difficult to approximate as they vary significantly according to the nature of the content (illustrations, colour, graphics can all impact on costs), journal extent and frequency, distribution medium and circulation.

The variation in the data points collected in the existing evidence base (see chart below) indicates that a 'broad-brush' approach to the cost profile of scholarly journal publishing probably oversimplifies the picture too much to be useful.

Journal publishing cost element	**Data ranges**
Content creation costs	16%-54%
Overheads	11%-55%
Manufacturing, printing and paper	8%-40%
Distribution and fulfilment	3%-17%

The gaps

We have some useful base data on costs of the current system of journal publishing. There are gaps and weaknesses, but, in the area of supply-side economics, many of these are in fact capable of being overcome through further research.

Very little up-to-date evidence is available regarding the cost (in terms of time or money) of the researcher community's contribution as authors, editors or peer reviewers to the current scholarly journals publishing process.

Evidence relating to the costs of launching new journals is also lacking.

The evidence base relating to publisher costs is partial in terms of its coverage of the publishing community (commercial and university presses in particular are under-represented) and it is not segmented sufficiently in terms of processes to address the particular questions being reviewed in this area of the current study.

The evidence of diversity in journal publishing economics provided by existing sources suggests that this issue will be most productively investigated at the journal level, and that data giving ranges of costs by discipline, frequency, extent and circulation will most accurately reflect the true complexity of supply-side costs.

Area 3: Usage

Key questions

This review of journal usage focuses on evidence relating to the use of journals by both readers and contributors. Key questions identified were:

1. How much use is made of journal articles?
2. What is the split between leading journals and the rest?
3. What is the extent of unmet demand for journals?
4. What is the potential for this demand?
5. What are the barriers to this demand being met?

The evidence

The evidence base on scholarly journal usage shows a good deal of consensus on key issues. However it is incomplete from a couple of perspectives. Firstly, the primary data focus on authors and users in higher education – corporate and other end-users are under-represented. Furthermore, much of the available evidence is based on user surveys rather than actual usage data (i.e. it records what users *say* they do, but we can not cross-validate this with actual *behaviour*). Valuable work has been undertaken in analysing transaction logs for electronic journal collections, but little of the data is currently in the public domain.

Readers value journals and particularly welcome the flexible access afforded by electronic content (although printing is still common practice for reading purposes). Journal reputation (and the peer review process underlying this) continues to be valued by readers and authors alike. Maintaining the peer review system is seen as important by all constituents of the user community, regardless of the publishing model used.

Surveys record that researchers can experience access problems (e.g. the library not stocking a particular journal), but these are not reported to be a major obstacle to research productivity overall. There is evidence that electronic resources are improving access. Interdisciplinary researchers are highlighted as facing the most difficulty in both accessing journals and in finding publishing outlets.

The gaps

Publicly available hard data on journal usage at the article level (in either hard or electronic copy) constitute a key gap. For example, it is possible that data logs showing unsuccessful requests for journal access would provide useful quantitative evidence of unmet demand (by readers).

Increasing electronic access to journals facilitates the collection of actual usage data and this resource will be critical to cross-validate and enrich data collected through survey methods.

There is no evidence relating to the precise nature of the use made of the information in journals that are accessed. We have no detailed view as to how access to journals 'adds value' to researchers' endeavours and productivity. This, in the context of emerging alternative modes of scholarly communication, is a significant gap to be filled by appropriate research.

Area 4: Citations, Impact Factors and their role

Key questions

The three questions considered in this area of the report are:

1. Are traditional (i.e., subscription-based) journals more likely to be cited than OA journals, and as a result, do impact factors (IF) vary between traditional and OA journals?
2. What impact do citations and IFs have on research funding?
3. How are IFs affected by the increasing trend to collaborative research and multi-location research?

The evidence

This is an area in which much research has been carried out, but most of it has been on specific subject areas or titles, making it difficult to generalise. Furthermore, any study of variance in article impact in OA environments versus subscription environments faces a key methodological challenge in that a given article cannot be OA and non-OA at the same time and, therefore, an exact like-for-like comparison of research impact over the same time period is not possible.

Much of the research that has been undertaken in this area has been on self-archived articles as opposed to articles that have appeared in OA journals; it is important to distinguish carefully between the two when interpreting the findings.

There is some consistency in results that show more citations for articles self-archived in repositories as distinct from the same or similar articles available in a subscription journal (although there have also been a few contradictory results). Overall, deposit of articles in open access repositories seems to be associated with both a larger number of citations, and earlier citations for the items deposited.

The reasons for this, however, have not been clearly established - there are many factors that influence citation rates, including the reputation of the author, the subject-matter of the article, the self-citation rate, and, of course, how important or influential the repository is in its own right. The little existing evidence suggests that a possible reason for increased citation counts is not that the materials were free, or that they appeared more rapidly, but that authors put their *best* work into OA format. This research was limited to one discipline, however, and more extensive evidence is required to validate this finding.

There is less consistent evidence relating to IF advantage for OA journals over toll-access journals, although OA articles in a non-randomised study in a hybrid journal have been recorded as achieving higher citation counts (over the same time) than subscription-access articles in the same journal.

With reference to any impact of citations and IFs on research funding, anecdotal evidence suggests that some funding agencies use citation counts as part of the assessment procedure when allocating research funds, but the number of publications confirming this approach is tiny. There are more data available on the reverse scenario, i.e., the impact of funding or source of funds on citation counts. We conclude that the two are linked, but that only rarely are citation counts a key factor in funding decision-making.

In relation to any link between citations and collaborative research, there is some scattered, but consistent evidence in various bibliometric studies that multi-institutional or multi-nationally authored papers are more frequently cited than papers that come from a single institute. However, such increases in citation counts that do occur with regard to co-operation are likely to be a purely numerical artefact of the greater number of authors and can be explained by self-citation.

The gaps

Although quite a lot of evidence has been collected regarding the quantitative effect of OA on citation counts (whether in the form of OA journals or as self-archived articles), much of it is scattered, uses inconsistent methods and covers different subject areas.

Consistent longitudinal data over a period of years to measure IF trends in a representative range of journals would fill this gap, e.g., studying a range of journals that were toll-access and went OA (or *vice versa*).

In the short-term, more data in different disciplines measuring the impact on citation counts of articles in hybrid journals or articles that are available in both forms *versus* articles that are only available in one of the forms will improve the evidence base.

Qualitative factors should not be ignored, however, in either conducting or interpreting research on this issue. The existing sources rarely take into account the full range of factors that can affect citation counts. These are challenging data to collect, but not unprecedented.

Little or no research has been carried out on research funders' approaches to citations as a metric for decisions. A comprehensive review of policy statements coupled with a primary survey of funding agencies would be required to collect base data on this issue.

The current evidence base does not support the premise that collaborative research leads to a meaningful increase in citation counts (and, by implication, greater research 'success') once self-citation by a larger number of authors has been taken into account. It is clear that any new research seeking to validate this finding through impact measurement must involve corrections for self-citations.

Area 5: Disciplinary differences

Key questions

This section reviews the evidence in two areas:

1. Is there any difference in the way researchers active in different disciplines use journals as readers?
2. Is there any difference in the publishing habits of different disciplines?

Findings in Area 5 are closely related to issues in Area 3 of this report and there is a high degree of overlap in the major sources consulted.

The evidence

Surveys of authors and readers (again mostly in higher education environments rather than industry) provide the richest data on these issues, but the focus, population samples, and methods of individual studies, together with the form and content of survey questions, obviously differ so direct comparison of results can only be tentative.

The survey evidence suggests that journal articles are most important in the sciences and social sciences, but that books are more important in the arts and humanities. There is some evidence of a trend towards greater convergence in the use of research resources between disciplines, but this finding needs further validation.

All researchers appear to have similar levels of access to the journal materials they need. The issue of ease of access to journals shows little meaningful variation by discipline – around 50% of all researchers, regardless of discipline, experience problems.

Faced with such barriers, only around a quarter of researchers encountering difficulties (both arts and sciences) never try to obtain articles via another method. Inter-library loan is the most popular alternative route in both the arts and the sciences. Those in the sciences are more likely to also turn directly to article authors for help than their counterparts in the arts.

All authors, irrespective of discipline, claim that career advancement and peer-to-peer communication are the most important reasons for publishing.

There is a clear consensus across sources on which factors influence an author's choice of journal in which to publish; impact factor, reputation and peer review were reported to be of primary importance to all authors, again irrespective of discipline.

Publication delays were identified as an obstacle impeding the publishing process in 1999 by nearly 50% of researchers in all disciplines, but most expected or hoped that electronic publishing with rapid peer review would remove this hurdle. We have no up-to-date evidence to validate whether this has in fact proved to be the case. Researchers in all disciplines anticipate that electronic dissemination of research will be increasingly important through to 2015.

The gaps

As in Area 3, analysis of transaction log data, recording usage activity (segmented by discipline), would add significant value and rigour to the information available in this area. The survey data we do have access to focuses quite narrowly on attitudes and preferences in relation to journals rather than exploring how journals are actually used in the context of researchers' professional lives.

The evidence we are able to draw on points to disciplinary convergence rather than divergence in researcher usage of and attitudes towards journals, whether from a reader or author perspective.

As highlighted in Area 3, a more large-scale exploration of the nature of discipline-specific journal usage – and most importantly the evolving role and value of journal articles alongside alternative modes of scholarly communication - would bring more depth to an area in which our understanding is superficial.

Area 6: Cost and impact of alternative formal dissemination models

Key questions

The key questions to be addressed in Area 6 are specified as:

1. What are the costs involved in publishing open access journals? To what extent are these different from those of publishing conventional journals?
2. What is the impact of digital repositories, institutional or thematic, on the economics of journal publishing?

The evidence

Sources in this area are diverse in focus, and more often than not provide glimpses of only limited areas of the topics under examination (as in, for example, close scrutiny of a small handful of journal titles, or examination of a narrow range of disciplines). The paucity of sources means that establishing evidence-based causal relationships in key areas cannot currently be done and, similarly, extrapolation from restricted samples to wider communities is currently not possible.

On the issue of costs, any evaluation of the impact of alternative models presupposes an understanding of existing models. However, the evidence about the costs of the traditional journal publishing process (described in Area 2 of this study) does not provide solid comparators against which OA costs can be set.

The focus of the current debate about the relative costs of an OA publishing system versus publishing conventional journals seems to be gradually changing. Our evidence review suggests a degree of acceptance that many of the components of cost are common to both principal models and can therefore be cancelled out in the 'equation' – for example, 'first copy' costs at the beginning of the cycle, and server and software costs at the other end.

The research base around 'costs' appears to be getting broader in scope. Work has now been undertaken on how new OA journal publishing models such as 'author pays', will be funded – and by whom. There is evidence suggesting that a straightforward institutionally-based solution would potentially be inequitable, concentrating the cost on a relatively small number of research-intensive institutions.

In 2005, the average number of items held in repositories was estimated to be 'a few hundred' (with the exception of the Netherlands and possibly the USA, for which details were not available). These items tend to be very diverse in nature, including to a very considerable extent teaching material as well as research contributions. An obvious exception is the long-established thematic repository, arXiv in the field of high-energy physics and related areas.

Whilst some evidence does suggest that these repositories are an important new factor in the journal cancellation decision process, and one which is growing in significance, there is no research reporting actual or even intended journal subscription cancellation as a consequence of the growth of OA self-archived repositories.

Subscriptions are reported to have been declining over a period of 10+ years, but for a number of reasons. Proving or disproving a link between availability in self-archived repositories and cancellations will be difficult without long and rigorous research. In this connection, the outcome of research recently announced by the Research Councils UK (RCUK),[1] with the co-operation of Macmillan, Blackwell and Elsevier, will be eagerly awaited, even though a report is not due until late 2008.

The gaps

On the issue of costs, the critical gap will be for a wide-ranging study on the funding and money flow implications of new publishing paradigms, particularly if the author-pays model becomes well established.

There is no evidence as yet to demonstrate any relationship (or lack of relationship) between subscription cancellations and repositories. Work in this field would need sufficient, representative and balanced samples, and the collaboration of all stakeholders, including especially research institutions and publishers. Any such study will need to be maintained over a fairly extended period, with regular reports, since it seems likely that the position could change with time if the contents of self-archiving repositories become progressively more comprehensive.

Similarly, more carefully conceived work on the impact of both OA journals and self-archiving on the quality of research communications, especially on the peer review system, will be required.

[1] http://www.rcuk.ac.uk/access/2006statement.pdf

Introduction

Scholarly journal publishing has been thrust into the public spotlight in recent years. The Competition Commission's investigation into the proposed merger of Reed Elsevier plc and Harcourt General, Inc in 2001[2] was the first of a number of very public attempts to dissect the particular dynamics of research publishing, a process which has most recently brought forth the European Commission-funded study on the economic and technical evolution of the scientific publication markets in Europe, published in January 2006[3]. Unfortunately, public scrutiny does not appear to have progressed or clarified the debate about what we do or do not know for certain about journal publishing.

Market studies measuring English-language publisher revenues value the scientific, technical and medical (STM) journals publishing market at around $5 billion globally in 2004[4]. STM disciplines cover a significant sub-set of the total universe of scholarly journals, but exclude important areas such as social sciences, arts and humanities subjects. Unfortunately we must conclude that even this figure is at best a partial estimate of the total market.

Compared with other sectors of professional publishing such as educational publishing (valued at $19.4 billion[5] in 2004) or global legal, tax and regulatory publishing (valued at $9.5 billion[6] in 2004), the scholarly journals segment (even assuming the true value will in actual fact exceed the $5 billion under-estimate) is a relatively small niche of the overall publishing industry. However, in the context of its role as the primary dissemination channel and public record of new research and development effort globally, the journal publishing process is both important and high profile.

In 2003, the UK's gross domestic expenditure on research and development (R&D) was £20,825 million[7]. Increasing this research and development expenditure is a focus of government attention as part of its Ten Year Science & Innovation Investment Framework outlined in 2004. The objective of the programme is to make the UK one of the best places in the world for science, research and innovation, and increasing R&D expenditure as a proportion of national income from 1.9% (2004) to 2.5% by 2014 is a key target. With R&D expenditure targeted to increase in the future, the means by which the outcomes of this investment are disseminated and accessed globally in the form of newly published research will necessarily come under continuing scrutiny. This is neither surprising nor inappropriate, given that more than a third of R&D in the UK is funded from the public purse through government investment and a similar proportion of spending on research publications is also publicly funded via academic and government library purchasing.

---

[2] Competition Commission, July 2001. Reed Elsevier plc and Harcourt General, Inc, *A Report on the Proposed Merger*, http://www.competitioncommission.org.uk/rep_pub/reports/2001/fulltext/457c1.pdf

[3] Dewatripont, M, Ginsburgh, V, *et al*. for the European Commission, D-G Research, *Study on the economic and technical evolution of the scientific publication markets in Europe*, January 2006. http://ec.europa.eu/research/science-society/pdf/scientific-publication-study_en.pdf

[4] $5 billion is the geometric mean of the EPS and Simba estimates for the global STM English-language journals market. This estimate covers a sub-set – if a substantial one - of the total potential journals market (it does not take into account all disciplines or foreign language publisher revenues). EPS Market Monitor, July 2005. *Scientific, technical & medical (STM) information: market trends and industry performance*. Simba Information, 2004. *Global STM market analysis & forecast*.

[5] EPS Market Monitor, January 2005. *Education and Training: market trends and industry performance*.

[6] EPS Market Monitor, May 2005. *Legal, Tax and Regulatory (LTR) Information: market trends and industry performance*.

[7] The Office of Science and Innovation, (DTI), *Science, engineering and technology indicators:* http://www.dti.gov.uk/science/science-funding/set-stats/gross-domestic-exp/index.html

Much of the recent research examining scholarly journal publishing has tracked a market in transition. Like the rest of the publishing industry, this segment has undergone its own revolution over the past 10 or so years, driven by the new dynamics of networked publishing. One might have expected the erudite field of primary research in academic disciplines to have been a laggard in the digital publishing revolution, but the evidence suggests quite the opposite. Uptake of online access to content has been far reaching and journals in either purely electronic or parallel print/electronic format account for over 60% of serial spending in the UK's higher education libraries[8]. New ways to access, search, filter and link digital content have already transformed the researcher's information gathering and analysing activity and experts predict that navigational, community and workflow solutions will be increasingly valued by researchers in the future. With developments such as blogs, podcasts and mobile devices still gathering pace, it is clear that this process of transformation at the researcher's desktop has only just begun.

The original research article is the key commodity which drives this marketplace and the efficiency of the system which creates and distributes this commodity is key. Alternative models for journal publishing are already being tried and tested in the marketplace and this has fuelled debate and questions from policy makers, research funders and market participants alike.

It is in this context that the Research Information Network (RIN), Research Councils UK (RCUK) and the Department of Trade and Industry (DTI) have jointly commissioned and funded this study. The objective is to audit and evaluate the available evidence relating to scholarly journals publishing, and furthermore to distil what we do and do not currently know for certain about this market niche. Thus, the current study provides a snapshot of the evidence base about the scholarly journal publishing process from a UK perspective and seeks to provide a basis of shared knowledge which policy-makers and research funders can draw on to consider if and how the system might develop, to ensure that the research publishing process delivers best value and best practice.

Our review of the primary literature relating to scholarly journal publishing has revealed data sources which range from in-depth facts and figures in isolated market niches to data which describe the market at so general a level as to risk oversimplifying it. Further it has been problematic to identify a 'UK perspective' given that UK authors publish and are read internationally, UK libraries and readers make use of journals from around the world, and UK publishers publish international authors and sell to international subscribers (and may, in some cases, be part of an international group).

We have sought to highlight specific major gaps in the evidence base and hope that this analysis will provide a clearer picture of what we *do* and *do not* know about the scholarly journal marketplace and will act as a roadmap for future research in this critical and dynamic field of publishing.

Methodology

The data presented in the study have been collected via an extensive literature review and have been validated and tested at critical points through consultation with experts in the field. An expert panel consisting of representatives from the academic, librarian, research funder and publisher communities was appointed to

[8] Creaser, C, Maynard, S, White, S (November 2005). *LISU Annual Library Statistics*.

the project. The panel reviewed research outputs during each phase of the project and met as a group to discuss feedback and amendments.

The project followed a three-stage process:

Phase 1: Scoping the study and identifying priority areas of interest;
Phase 2: Identification and evaluation of sources;
Phase 3: Data collection, triangulation and identification of gaps.

Scoping the study

During the scoping phase of the project, six priority areas of interest – and a number of key issues within each - were identified for review (as specified in Section 1.2 of the original project specification):

1. Journal market value and volume;
2. Journal supply-side economics;
3. Usage;
4. Citations, impact factors and their role;
5. Disciplinary differences;
6. Cost and impact of alternative formal dissemination models (e.g. open access and digital repositories).

Initial literature search to identify sources

A literature search was undertaken to identify relevant sources in each of these six topic areas. Following an initial trawl, the list was circulated to both the Expert Panel and a range of other industry experts to elicit additional suggestions and to ensure that there were no important omissions.

Filtering the list of key sources

Each source on the preliminary list was then carefully reviewed. Many sources were disregarded at this stage as not applicable to the study. Our objective was to identify sources which provide robust *primary* data and which, ideally, permit some kind of UK-level analysis or extrapolation. In some areas where no UK-based data exist or where geographic distinctions are less relevant to market characteristics, we have referenced studies with an international scope. The unique primary surveys of scientists undertaken by Tenopir and King in the US are a good example. As noted above, the underlying market dynamics of scholarly journals are, by nature, global: UK authors publish and are read internationally, UK libraries and readers make use of journals from around the world, and UK publishers publish international authors and sell to international subscribers (and, indeed, may in some cases be part of an international group). Some secondary sources were also deemed relevant when they built substantively upon reliable primary data. In general, we have discounted data published before 2000 as too old to be relevant, but again, in areas where data are sparse or key sources have not been updated, there are exceptions and these are clearly cited in the study.

Evaluation of the sources

The data reported below are based on sources which have been carefully evaluated in terms of their applicability to the study. One of the key objectives of the project was to develop a method of source evaluation that was as objective as possible and would be applied consistently throughout the data collection activity. Appendix 1 details the method and criteria used for evaluating each

source and provides definitions of the criteria and the ranking system which have been applied.

Each major source referenced in the report has been ranked in one of three categories:

- A= Data applicable to our research
- B= Selected data applicable to our research
- C= Data with very limited applicability to our research

Consideration is also given to the relative relevance of the source to the UK market. It is important that this evaluation process is transparent to those using this study. To this end a Source Evaluation Sheet has been completed for each major source referenced in the study and these have been published in Appendix 1.

Data analysis and 'triangulation'

In the third and final phase of the study we have analysed and synthesised applicable data identified in the major sources and present this evidence in a written report. The report discusses each of the key sources identified, analyses the available data and identifies gaps in the evidence base. It is important to point out that, in many areas, the available data points are either too specific or too general to permit useful cross-analysis or triangulation. These areas have been clearly identified in the text.

As a final step, each chapter of the draft report findings was reviewed and validated by a different market expert in the scholarly publishing field. Feedback was gathered from each via a telephone or face-to-face interview and comments were integrated into the final report. This report was then submitted to the appointed Expert Panel for final review before publication.

Structure of the report

This report is organised into six sections, each of which addresses one area of the journals publishing marketplace selected for investigation (as specified in Section 1.2 of the original project specification and listed above).

Each of these six sections follows the same basic structure and is divided into four sub-sections:

1. *Scope and definitions:* this defines what will be included (and excluded) from scope in each section and offers definitions of key terms as they are used and understood in the context of this study;
2. *Key sources:* in this section we discuss each of the key sources identified and discuss their relative strengths and weaknesses as a data source;
3. *Analysis/data validation:* this section is concerned both to validate and to triangulate key data points gleaned from the evidence base described under 'Key Sources'. Where possible, we have either cross-compared data sources in order to validate robustness or sought to triangulate available data-points with other sources to fill key gaps or provide proxies or benchmarks. However, for validation of data to be possible, it is essential to have established benchmarks, or at least to have datasets which have the same or similar terms of reference, with any non-overlapping areas clearly identified and their boundaries properly defined. This is not the case in several of the Areas discussed in this study, where such data as do

exist are diverse in focus, and more often than not provide glimpses only of limited areas of the topics under examination (as in, for example, close scrutiny of a small handful of journal titles, or examination of a narrow range of disciplines).

- *Gaps in the data:* discussion in the preceding text and the detailed evaluation sheets in Appendix 1 have sought to highlight the degree to which propositions are underpinned by evidence or not. Here we aim to summarise the key gaps in the evidence base and also distinguish between those gaps that can be filled through further research (what is capable of being proved?) and those issues which remain uncertain (what cannot be known for certain).

Two appendices accompany this report:

Appendix 1: Source Evaluation Sheets: this Appendix includes a Source Evaluation Sheet for each of the major sources reviewed in this report. The source evaluations are organised alphabetically by author name and numbered sequentially. Sources are further grouped by topic (i.e. areas 1-6 as listed above) and numbered according to both topic and sequence (i.e. Source 1.1=Area 1, Source 1, and so on). The reference for each source (e.g. 'Source 1.1') is quoted in the main body of the report, enabling easy cross-reference from the text to the Source Evaluation Sheets as required. The first section of Appendix 1 documents the methodology and criteria used for evaluating each source and provides definitions of the criteria and the ranking system which have been applied.

Appendix 2: Project Specification: this document outlines the original scope, aims and objectives of the study.

Guide to using this report

We would recommend that readers have the relevant Source Evaluation Sheets for each area to hand when reading the main body of the report. Cross-referencing to the Source Evaluation Sheets will provide a useful recap and reminder of the different sources as they are mentioned in the analysis.

Acknowledgement of experts consulted

We should like to take this opportunity to thank all the experts who have contributed to this report for their invaluable input and support.

Expert Panel Appointed to the Project

Professor Jeffrey Aronson, British Pharmacological Society
Paul Ayris, Director of Library Services, UCL
Juan Bicarregui, Head of the e-Information Technology Sector, CCLRC
Bob Campbell, President, Blackwell Publishing Ltd.
Sally Morris, Chief Executive, ALPSP
David Prosser, Director, SPARC Europe
Stephen Pinfield, Deputy Chief Information Officer, University of Nottingham

Expert Witnesses – Report Validation

Mayur Amin, Director of Research and Publisher Relations, Elsevier
Cliff Morgan, Vice President, Planning and Development Director, John Wiley & Sons Ltd
Professor John Wood, Chief Executive, CCLRC/Head of JISC Scholarly Communications Group
Professor Keith Jeffery, Director, Information Technology and Head of Business and Information Technology Department, CCLRC
Diana Garnham, Chief Executive, The Science Council
Claire Jenkins, Director of Library Services, Imperial College

Experts Consulted to Validate the List of Key Sources

Robert Long, Sales and Marketing Director, John Wiley & Sons Ltd.
Stevan Harnad, School of Electronics and Computer Science, University of Southampton
Peter Suber, Open Access Project Director at Public Knowledge, Senior Researcher at SPARC, author of the Open Access News blog and SPARC Open Access Newsletter
David Worlock, Chairman, Electronic Publishing Services Ltd

Area 1: Journal market volume and value

Executive summary

Key questions

The particular issues identified for review in this section are:

1. Industry volume and value
2. Split between commercial and society publishers, volume and value
3. Acquisition and cancellation trends
4. Split between print and electronic revenues
5. Split of publisher revenue between corporate, academic and personal subscribers
6. Value of PPV article sales

The evidence

Volume

There are estimated to be around 20,000-25,000 peer-reviewed scholarly journals actively being published worldwide. The volume of journals has grown consistently at a compound annual growth rate of 3-4% over the past one hundred years.

An estimated 60% of all journals are published online (most are also parallel published in print). This figure is reported to be as high as 90% for English-language journals emanating from US/UK-located publishers. Around 10% of peer-reviewed scholarly journals are published under some form of open access model.

There are estimated to be at least 200,000 full time researchers working across industry, government and higher education in the UK. Global sources suggest there may be 5.5 million researchers globally.

Value

Serial subscriptions spend by UK higher education institutions (of which peer-reviewed scholarly journals are a sub-set of unknown size) totalled £94.5 million in 2003-2004. The number of current serial subscriptions has doubled in higher education institutions over the past ten years. Much of this growth in volume is thought to be attributable to 'bundled' or packaged deals offered by publishers.

Published sources estimate publisher revenues of circa $5 billion for English-language journals in scientific, medical and technical disciplines in 2004. It is likely that this is a significant under-estimate of the total peer-reviewed journal market (it excludes non-English language journals, journals in disciplines outside of STM such as social sciences, the arts and humanities and is necessarily based on 'best estimates' of journal revenues derived from figures in published accounts).

The customer profile will vary from journal to journal depending on business model and its disciplinary focus. Overall it is likely that academic customers account for the largest proportion of these revenues (c30-40% of revenues), with significant numbers of corporate customers (c20-30% of revenues), and

government, individual and professional subscribers (e.g. healthcare professionals) and advertising clients (such as pharmaceutical companies) accounting for the balance.

The gaps

The somewhat dislocated story above reveals much about the gaps which undermine the existing baseline data about scholarly journals.

In terms of the market constituents, we have particularly poor data about buyers of scholarly journals beyond the higher education customers. Even basic data such as the numbers of these types of special libraries and information centres which exist in the UK, are not published.

Further work on sizing and segmenting the publisher market would also be valuable – we have indicated how sensible estimates can potentially be derived from available sources such as trade associations.

The general nature of published reports regarding the size and structure of the journals publishing market hinders our interpretation of important trends in this marketplace. Studies gathering data relevant to the STM publishing market are too broad in product focus and too narrow in terms of disciplinary coverage to deliver in-depth data relevant the market defined as the subject of this report.

Deeper analyses of the value of scholarly journal publishing (such as sales to different types of customers, sales in different formats and the balance of different revenue streams) are not possible.

More survey data (anonymised and aggregated) directly sourced from the publishers themselves and tracked over time would go a long way to filling the gaps in the evidence base on market size and structure.

Introduction

The key purpose of this section of the report is to identify evidence that accurately describes the size and structure of the journals publishing market in terms of both volume (e.g. numbers of participants, numbers of journals, numbers of articles) and value (demand-side/buyer-side spending on journals and supply-side/publisher revenues generated through journal sales).

Looking in more detail at the demand side of this market (i.e. the researchers and libraries who purchase and use journal products), we also explore evidence relating to acquisition and cancellation trends. On the supply side (the publishing segment), evidence is sought which describes the split of revenues between different types of publishers (e.g. non-profit distributing versus profit distributing publishers), different customer segments (corporate, academic, and personal subscribers), and different journal formats (print and electronic).

In summary the issues specified for coverage in Area 1 of the report are:

1. Industry volume and value;
2. Split between commercial and non-profit distributing publishers, volume and value;
3. Acquisition and cancellation trends;
4. Split between print and electronic revenues;
5. Split of publisher revenue between corporate, academic and personal subscribers;
6. Value of PPV article sales.

Scope and definitions

This report is concerned with the market for scholarly, peer-reviewed journals in all disciplines and is principally concerned with examining the publishing of primary research articles. This niche publishing market mainly comprises journals that present original research articles, but also includes some secondary titles (which publish reviews, summaries and abstracts of primary journal content). These other types of journals are included in sources of market size estimates and journal counts, primarily because it is virtually impossible to separate them. Equally, review articles are often a composite part of many primary journals and sit alongside primary articles.

Scholarly journals have both a global author and customer base. In addition, many of the larger publishers operate internationally. As a result most meaningful analyses of the market are also, necessarily, global. However, where possible and appropriate, we have reported information concerned specifically with the journals publishing market in the United Kingdom.

For the purposes of this report, we define the 'demand side' of the journals market as:

- Individual researchers who read and use journal content in their work or study (whether in an academic, government or corporate environment); and
- Libraries in academic, research, public, and corporate organisations who are the main purchasers of scholarly journals on behalf of individual researchers.

We define the supply side of the journals market as:

- Individual researchers who author the journal content (while all authors are also readers, not all readers are authors, so the distinction is an important one);
- Publishers who manage the creation, production and distribution of full-text journal content. These can be broadly divided into two categories: profit distributing (i.e. commercial) publishers and non-profit distributing publishers such learned societies, professional associations, intergovernmental organisations (such as OECD, WHO, ILO), research foundations, charities etc. and university presses; and
- Secondary publishers, such as content aggregators and abstracting and indexing services, which provide end users with access to content from journal publishers, primarily in electronic format.

Key sources

Several sources have been identified as contributing key data on these issues.

Title and article volumes and trends

There are surprisingly few good basic data relating to the volume and structure of scholarly journal publishing output. Some basic top-level numbers can be derived from major reference sources, but in-depth market data that describe and segment the universe of journal titles according to discipline, frequency, extent, type of publisher and other characteristics are lacking.

The main sources of data on title and article output and trends are the bibliographic database Ulrich's Periodicals Directory[9] (Source 1.4) and Thomson Scientific's Journal Citation Reports and Indexes[10] (Source 1.13). Ulrich's is a standard bibliographic reference work used by libraries globally. The directory lists serial publications that are published regularly or irregularly and are circulated free of charge or by paid subscription. First published in 1932, it now lists more than 260,000 print and electronic periodicals of all types: academic and scholarly journals, consumer and trade magazines, newspapers, newsletters and bulletins. Listings for serials titles are free, but voluntary. While it is unlikely that the directory is 100 per cent comprehensive, it will certainly include all titles of note and is the most inclusive and authoritative source available. It is widely cited in the literature relating to the scholarly journals market as the key source for sizing annual journal title output and growth for example, see Mabe and Amin[11] (Source 1.9), who provide an analysis of historical journal title growth trends.

While the overall scope of Ulrich's is too broad to be relevant to this study, the online version of the database[12] can be searched using more than twenty filters, including categories for 'active refereed scholarly journals', 'online journals' and 'open access journals'. Country of publication is also included as a search term, although the multi-national character of journal publishing in many of the larger firms undermines the usefulness of this parameter. It should also be noted that there is a natural time-lag in new journal titles entering the Ulrich's database (there is consistently a drop in title growth in the most recent 2-3 years) and journals are continually being added retrospectively as they come on to the Ulrich's editorial radar – so journal universe counts will change over time. It is also true that Ulrich's is unable promptly to register cessation of publication. Unfortunately, the database does not distinguish between types of publisher nor does it assign imprints to their ultimate publisher owners, so is rather less useful in terms of sizing and segmenting the supply side of the market.

The primary purpose of Thomson Scientific's Journal Citation Reports (JCR) and Indexes (Source 1.13) is to report article citation counts that are used to measure journal impact factors (IFs). More detailed discussion of IFs is contained in Area 4 of this report. As a by-product of this activity, the Journal Citation Reports are also widely used in the literature to quantify and analyse research community article output and growth. However, it is critical to note that the citation reports are not all-inclusive; journals are only included on the basis of a number of

[9] CSA, *Ulrich's Periodicals Directory*.
[10] Thomson Scientific (ISI). *Journal Citation Reports and Indexes*.
[11] Mabe M & Amin M, (2001). "Growth Dynamics of Scholarly and Scientific Journals" in *Scientometrics* 51 (1).
[12] Ulrich'sweb: http://www.Ulrich'sweb.com

selection criteria as defined and implemented by Thomson Scientific's editors[13]. The reports cover 8,700 peer-reviewed journals in approximately 200 disciplines (the Science Edition covers over 5,900 science journals; the Social Sciences Edition covers over 1,700 journals and the Arts and Humanities Edition covers 1,130 journals). As Thomson Scientific have modified and expanded the selection criteria over time, any longitudinal analyses of JCR data should be interpreted cautiously.

Two surveys undertaken by Cox[14] (Source 1.3) on behalf of ALPSP provide useful primary data on journal volume and growth trends to cross-compare with secondary analyses of Ulrich's Directory (Source 1.4) and Thomson Scientific's (Source 1.13) citation reports. These surveys seek primarily to report trends in publishers' practices and policies, with particular reference to online publishing. The second survey draws on a sample of 174 valid responses from scholarly publishers (25 per cent commercial publishers versus 75 per cent not-for-profit publishers; 40 per cent of total respondents are UK-based organisations). A balanced range of disciplines and size of publisher is represented, including 13 larger publishers with >100 journal titles in their portfolio. Cox analyses the number of journal titles launched, discontinued, merged and transferred or sold over a five-year period (2001-2005). Results are split out by small, medium and large publishers and by commercial versus not-for-profit publishers. Unfortunately, the survey collected size of title output per publisher in bands rather than actual totals, and hence growth trends in title output across the sample cannot be readily calculated, although clearly some estimates can be derived. Journal title growth in terms of disciplinary differences is not analysed in the report.

Volume and value of demand side (customer) market segments

The researcher segment

The Office of National Statistics and the DTI's Office of Science and Innovation[15] (OSI) collate a range of statistical indicators relating to R&D funding and expenditure in the UK's science, engineering and technology sectors. It should be noted that these statistics do not cover researchers active in all segments (arts, humanities and social sciences are clearly outside of OSI scope). Longitudinal data relating to the overall population of the research and development community in the UK outside the higher education sector through to 2003-2004 are provided, along with a breakdown by the type of institution to which they are affiliated or by whom they are employed (i.e. business enterprise, research council, government department, or private non-profit organisation) and by specific role (researchers versus technicians versus administration and other staff). The 2003-2004 estimates for the researcher population in the UK's science, engineering and technology sectors are 116,000 FTEs, of which nearly 90 per cent are employed in the business enterprise segment.

As indicated above, data relating specifically to researchers employed within the UK higher education segment are *not included* in this total. Data on academic staff volumes and roles are sourced separately from the Higher Education Statistics Agency[16] and are analysed by activity (teaching and research staff

[13] The Thomson Scientific Journal Selection Process (last updated January 2004): http://scientific.thomson.com/free/essays/selectionofmaterial/journalselection
[14] Cox J and Cox L (2003 and 2005). *Scholarly Publishing Practice – The ALPSP report on academic journal publishers' policies and practices in online publishing (First and Second Surveys)*
[15] Office of National Statistics & DTI Office of Science and Innovation. *SET Statistics – Science, engineering and technology indicators*: www.dti.gov.uk/science/science-funding/set-stats/index.html
[16] UK Higher Education Statistics Agency: www.hesa.ac.uk

versus research only versus teaching only versus neither teaching nor research), by career grade (professor, senior lecturer, researcher, etc.) and by employment status (full or part-time). HESA data suggest there were 101,000 active full-time active research/researcher and teaching staff employed in UK higher education institutions (HEIs) in 2004-2005 (part-time personnel would account for an additional 18,300). Further institutional and disciplinary breakdowns can be derived from the Research Assessment Exercise (2001),[17] although this information is now somewhat dated.

As we have noted above, the addressable market for scholarly journals tends to be global in nature; estimates of the global research community are regularly compiled by UNESCO,[18] which collects statistics directly from relevant national statistical offices as part of their Survey of Science and Technology Statistics. Their latest publicly available estimate (source data relating to 2002 or earlier) estimates a base of 5.5 million researchers worldwide.

It should also be noted that there is a distinction between the core active researcher segment and the wider journal reading community, which is likely to be larger. These additional readers will include practitioners (such as healthcare, legal or engineering professionals) and undergraduate students in universities, plus other more peripheral and infrequent readers. We have seen no robust evidence sizing this wider journal reader community. Internal research at Elsevier[19] derived from analysing global unique user counts for ScienceDirect suggests that the total global journal readership may be between 10-15 million.

Data quantifying the direct spend of individual researchers on accessing scholarly journal content is thin. Large-scale surveys of US researchers and scientists undertaken by Tenopir & King[20] (See Area 2, Sources 2.7, 2.8, 2.9, 2.10) across the 1970s, 1980s and 1990s document a fall in the average number of individual subscriptions to journals on the part of both university and non-university scientists and an increase in the volume of material accessed via a library facility.

The library segment

The constituents of the UK library universe may be broadly categorised as: public libraries, higher education libraries, national libraries, government libraries, special libraries/information centres, and school/FE college libraries.

In the context of journal purchasing, it is the higher education libraries, special libraries, national libraries, and government libraries which are the most significant (the first two in particular). Statistical data collection relating to the activities of all but the special libraries segment is relatively robust. Data relating to special libraries and information centres is a key gap in the evidence base. Even basic data, such as the number of these units which exist in the UK, are not published. These are libraries in private enterprises that are most likely intensive investors in R&D, e.g. companies in pharmaceutical and biotechnology, chemicals, oil and gas, engineering (e.g. aerospace, automotive, defence), electronics, telecommunications and IT sectors, plus other information intensive professions such as law firms. The Special Libraries Association (SLA) based in the USA is an international grouping of special library professionals dominated by North American members. The SLA has not published size or segmentation data

[17] UK Research Assessment Exercise (2001): www.hero.ac.uk/rae
[18] UNESCO, (2005). UNESCO Science Report: www.uis.unesco.org
[19] Telephone interview with Mayur Amin, Director, Research and Academic Relations, Elsevier (Oxford) on 28 July 2006.
[20] Tenopir & King, (December, 1998). Designing Electronic Journals with Thirty Years of Lessons from Print, *The Journal of Electronic Publishing*, Volume 4, Issue 2.

relating to the special libraries market. No comparable umbrella organisation has been identified in the UK, although some vertically organised associations, such as the Pharma Documentation Ring and the City Information Group, do exist.

The primary source of information relating to spending trends in HEIs is the Society of College, National and University Libraries (SCONUL)[21] (Source 1.12). SCONUL collects statistical information annually from its members relating to library usage, costs and spending, and details this by institution. SCONUL membership is open to all HEIs and most of the UK HE sector is covered in the statistics gathered. The latest available return (2003-2004) elicited an 82 per cent response rate covering >91 per cent of UK FTE students and 135 out of a possible 164 responding institutions. The key data of interest to the current study (and available by institution) include the overall breakdown of expenditure (staff vs. information provision vs. equipment vs. other operating costs), total serials expenditure, serials expenditure per FTE student, serials expenditure as a percentage of total information expenditure and number of subscriptions. One limitation of the data is that peer-reviewed scholarly journals are not specifically broken out from the general 'serials' umbrella category.

A further key source of evidence regarding the size, structure and spending of the UK library segment is the Library Information and Statistics Unit based at the University of Loughborough. Their Annual Library Statistics[22] (Source 1.7) provide a selection of key measures derived from SCONUL and other primary sources to present aggregated trending data on library spending over ten years (the latest data range from 1994-2004). Data available for national and government libraries are far less detailed than for the higher education segment.

A resource providing a detailed overview of international library spending on scholarly journals has not been identified. Summary statistics for academic libraries for a number of European countries are available from the European Commision's LibEcon[23] project although these data are now somewhat old (data go up to 2001). The Association of Research Libraries[24] (ARL) is a grouping of leading research libraries in North America which collects and publishes a range of longitudinal data from American Research Libraries. Data covered include collections, staffing, expenditures, library services, and library and university characteristics for the 112 ARL university libraries. Data relating to electronic content are available from 2004 onwards. Neither of these sources throws any light on the UK marketplace.

Acquisition and cancellation trends

Data from UK HEIs via SCONUL will be useful here in terms of tracking trends in institutional subscriptions to journals (e.g. numbers of current serial subscriptions and serials acquisition as a proportion of overall information spend). Once again, however, any data relating to 'serials' will overestimate our more niche target market of peer-reviewed, scholarly journals. There is little published evidence tracking subscription trends in the publishing community – this is commercially sensitive information and not in the public domain.

Similarly there is robust trending data relating to the growth or otherwise in individual subscriptions, although, as noted above, Tenopir and King (See Area 2) do document a fall in the average number of individual subscriptions to journals

[21] Society of College, National and University Libraries (SCONUL), *Annual Library Statistics*.
[22] Creaser, C, Maynard, S, White, S (November 2005). *LISU Annual Library Statistics 2005*.
[23] LibEcon web site: www.libecon.org
[24] Association of Research Libraries statistics and measurement program home page: www.arl.org/stats

in their research through the 1970s to 1990s[25]. Tenopir and King's data is drawn from a series of surveys throughout this period and covering a total of 13,500 scientists in both academic and corporate environments. Although the surveys are wholly US-based there is no obvious reason why the findings are not also applicable to the UK.

Also relevant here is Ware[26] (Source 1.16) who undertook a primary survey of 340 librarians to examine the issues that impact on journal cancellations and, in particular, to establish whether or not freely available pre- and post-print articles are a key factor in librarian cancellation decisions (see also, Area 6 of this report). The study collects data on who the decision maker is in libraries regarding cancellations and also describes the process. The three most important factors in journal cancellations are identified as: 1) faculty no longer require the journal title, 2) usage and 3) price (absolute price and price growth). The study offers useful contextual data about the relative and potential impact of new journal content access models on acquisitions/cancellations decision-making, but it provides no absolute figures relating to current cancellation behaviour.

Journal pricing

Data sources examining journal acquisition and cancellation trends necessarily also address the issue of journal pricing. Most of the analyses reviewed below use institutional subscription price levels as their base data point. However, many purchasers now opt for multi-journal packages or consortia-facilitated deals incorporating both print and electronic formats where 'bundled' pricing is the norm, all of which makes title-by-title price disaggregation very problematic. Further we have found no published primary evidence relating to either a) the scale of uptake of such deals or b) their overall impact on pricing trends.

Aggregated data on pricing trends for individual title subscription prices are found in LISU[27] (Source 1.7) and Van Orsdel & Born[28] (Source 1.15). The former source presents data collected in the Swets Serials Price Increase Reports. Swets merged with rival subscription agent Blackwell in 2001. Before this Blackwell had produced an annual International Periodical Pricing Survey. The new Swets report is compiled on a different basis from the Blackwell data making long-term trending problematic. However, Swets has provided LISU with retrospective data for the purposes of its Library Statistics Report. Hence, the data reported in their report are reliable and consistent over time.

Van Orsdel & Born[29] (Source 1.15) provide trend analysis of journal pricing data based on analysis of a sample of titles contained in the Thomson Scientific/ISI Citation Indexes (4,893 titles) and EBSCO Publishing's Academic Search Premier (2,759 titles). Price histories are obtained for the sample of titles from the EBSCO database and aggregated to provide a range of price-trend metrics for 2001-2005. This is an annual survey carried out for *Library Journal*. Key metrics provided are: average journal price by broad discipline, by niche subject, by country (of origin of content) and by region, plus annual increase and five-year 2001-2005 increases for these segmentations.

[25] Tenopir & King (December 1998). Designing Electronic Journals with Thirty Years of Lessons from Print, The Journal of Electronic Publishing, Volume 4, Issue 2.
[26] Ware, Mark, on behalf of ALPSP (March 2006). *Survey of librarians on factors in journal cancellation*.
[27] Ibid.
[28] Van Orsdel, LC & Born, K, (2005). *Library Journal*, Periodical Price Survey, April 15, 2005
[29] Ibid.

Data from this source can be cross-compared at the top level with data from the annual Swets Serial Price Increase Report. However, the subject and geographical classifications across these two studies differ. Equally, the Swets survey draws on a far broader sample (and definition of) periodicals, taking its data from titles listed in the Swets database (>67,000). The Van Orsdel & Born survey focuses only on a sub-set of titles listed in ISI Citation Index as a result is probably a more representative sample of peer-reviewed scholarly journals than the broader sample taken by Swets.

A more detailed collection of individual journal pricing data is available from the Journal Value Project (JVP) (Source 1.14) implemented by library personnel at the University of Wisconsin[30]. The project's aim is to develop a database that will collect journal cost and valuation data. Thus, its purpose is not to track pricing trends over time, but rather to attempt to quantify the 'value-for-money' of individual journal titles using a number of metrics (price per page, price per article, price per 1,000 characters and price per 1,000 characters divided by IF). The database identifies the individual publisher of each title and classifies their status as 'for-profit', 'not-for-profit,' or 'for-profit on behalf of society'. It is interesting to note the inclusion of this third category (whereby learned society publishers subcontract their journal publishing operation to a commercial publisher). This is a signifcant category of the marketplace (though once again not robustly quantified in the evidence base) which is often overlooked. The current JVP database includes data for 2001 (circa 3,000 titles) and 2004 (2,500 titles from the 2001 list for which the Univeristy of Wisconsin still maintained a subscription in 2004 *and* which are covered in Thomson Scientific's Citation Indexes). Coverage of the pricing survey is currently limited to STM titles and, by definition, includes only those titles to which the University of Wisconsin has access; it is envisaged that journals in the humanities and social sciences may be added in the future. The raw data are available in full in Excel, but unfortunately no secondary analysis of the data set is provided.

A broader, but similar database[31] has been compiled by Bergstrom & McAfee (Source 1.2) at the University of California's Journal Cost Effectiveness web site. Bergstrom & McAfee have databased a range of pricing data metrics for nearly 5,000 journals across a range of 17 broad discipline classifications. Raw data are sourced from Ulrich's (2004) and Thomson Scientific's Journal Performance Indicators (1998-2002) and so the impact on price per article resulting from changes in article volume *after* 2002 are not reflected in the data. Summary statistics derived from the database are made available online[32] and Davis[33] provides further secondary analyses of selected data.

Cross-comparison of these two data sets is certainly possible on a price per article basis and sub-segmented by discipline. This would be a useful, though substantial exercise in order to cross-validate and explore Davis' findings, which show for-profit publisher pricing at the top level to be significantly higher than not-for-profit journals, particularly for the very largest commercial publishers (in terms of both median cost per article and median cost per citation). Clearly, it would be helpful to undertake this analysis in more detail to identify, for example, trends at discipline level, costs at the per page level (article length can vary considerably by discipline for example) and to take account of the 'for-profit on

[30] University of Wisconsin, Madison Campus Libraries: 'Journal Value Project (JVP)', www.wendt.wisc.edu/projects/jvp/welcome.do
[31] Bergstrom, T & McAfee, P. Journal Cost Effectiveness Web Site: www.journalprices.com
[32] Bergstrom, T & McAfee, P. Journal Cost Effectiveness Summary Statistics: www.hss.caltech.edu/~mcafee/Journal/Summary.pdf
[33] Davis, P (Revised December 2005). Journal Cost Effectiveness Tables and Graphs: http://people.cornell.edu/pages/pmd8/

behalf of society' category which is not analysed separately in Davis' data. It would also be instructive to consider pricing differentials in the light of 1) journal circulation and 2) journal revenue model (i.e. which revenue streams other than subscriptions contribute to journal income?). Both of these factors clearly impact on publisher pricing policies, but are not taken into account in existing pricing analyses.

Davis[34] (Source 1.5) takes a different approach to assessing the relative value for money of journal expenditure in his attempt to cross-compare institutional costs of the current subscriber-pays journals system with the potential costs of an author-pays system for a sample of 113 institutions designated under the Association of Research Libraries (ARL). His spreadsheet relates the cost of subscribing to journals to the estimated article output of each institution. Of course, this is a purely academic exercise, as under the current subscription model there is no direct relation between an institution's research output and the cost of journal access. Thus, in the current context of examining demand-side expenditure trends on scholarly journals, the data tell us little other than reinforcing the fact that all those currently accessing the scholarly journal literature pay the same access fees regardless of their individual article input into the system. The main objective of Davis' model, however, is to relate each institution's individual contribution to the research knowledge base (i.e. article output by first author) to what it pays to gain access to that knowledge base (i.e. journal expenditure) in order to compare these results to a range of costs postulated in the producer-pays open access model. However, Davis' data rely on a number of estimated data points (including first author article output per institution) and can be treated as indicative at best. He finds that the institutional cost of funding the scholarly journals system for the vast majority of ARL institutions would be likely to be higher under a producer-pays model than current subscription fees (see also Area 6).

A further study comparing relative pricing/value metrics for a small sample of specific publishers was undertaken by LISU[35] on behalf of Oxford University Press. Data were collected in 2003 relating to 12 publishers (eight commercial players and four university presses) including OUP. The selection criteria for the survey did not seek to obtain a structured or representative sample of journal publishers. The sample was pre-selected by OUP to include only its main competitors. Thus, the results are indicative of the relative pricing/value positioning of these players to one another, but are not representative of the market as a whole and consequently are of rather less value in considering pricing trends generally. Following publication of the report, it came to light that some secondary titles (abstracting and indexing publications) had been included in the analysis where they should not have been and that the complex issue of change of ownership of certain titles over the period of review was not always accurately reflected in the data. As the data set has not been re-worked to take account of these inconsistencies, we have not included these findings in our analysis.

Volume and value of supply side market segments

Authors

We have found little primary research relating to the overall number of journal article authors in the UK. Clearly this will be a sub-set of the total research

[34] Davis, Philip and the Cornell University Library Task Force on Open Access Publishing. Dec 22, 2004. Calculating the Cost per Article in the Current Subscription Model – http://people.cornell.edu/pages/pmd8/

[35] White, S, & Creaser, C, (2004). *Scholarly Journal Pricing: Selected Trends and Comparisons*. Loughborough, LISU.

community described above (see: The researcher segment). Survey data from Tenopir & King (see Area 2; Source 2.7) identified a proportionate increase in article authorship by university scientists, as opposed to corporate or government-based scientists, between 1975 (62 per cent) and 1995 (75 per cent). In a later article (See Area 2; Source 2.9) the authors present data to describe how not all readers of articles are also authors, i.e. there is an extensive non-author reading community. Their US-based research indicated that, in 1995, scientists in US universities accounted for between 10-20 per cent of the total US researcher community, but accounted for 75 per cent of articles produced. At the same time, they accounted for <25 per cent of all reading. More recent work from Tenopir & King (Area 2; Source 2.10) proposes that about 15 per cent to 20 per cent of scientists in the United States have authored a refereed article. This estimate – and the asymmetry between authors and readers – is corroborated by work from Mabe and Amin[36] who estimate that, of the 5-6 million global researchers calculated by UNESCO, only around 1 million (circa 18 per cent) are unique repeat authors.

Publishers

We have not found robust attempts to size or segment the total population of either UK-based or global scholarly journal publishers in the existing literature. Official statistics from the Department of Trade & Industry[37] size the total UK publishing market at 8,000 organisations, of which UK-based journal publishers will be a sub-set. ALPSP and CAPP[38] (2002) estimated the population of UK-based journal publishers to be 'several hundred', but no hard evidence has been identified to confirm this.

In its submission of evidence to the House of Commons Science and Technology Committee (2004), The International Association of STM Publishers (an international trade body with around 100 publisher members) quotes a figure of 2,000 STM publishers worldwide. In a joint statement, ALPSP and STM say that together they represent around 300 non-profit and commercial publishers who are responsible for over 50 per cent of journal titles.

Careful further analysis (and deduplication) of the membership of the key publisher trade bodies in this field internationally would potentially yield useful data to substantiate these estimates and sub-segment the market more effectively between commercial, society and university press publishers (note: ALPSP has already undertaken some of the groundwork in this area internally, but has not published the outputs), e.g.:

- UK (The Council of Academic and Professional Publishers [part of the Publishers Association], The Periodical Publishers Association, the Independent Publishers Guild, The Scottish Publishers Association, The Welsh Books Council);
- USA (the Professional and Scholarly Publishing unit of the Association of American Publishers, the American Association of University Presses);
- International (the Association of Learned and Professional Society Publishers, the International Association of Scientific, Technical and Medical Publishers).

[36] Mabe M.A. and Amin M. (2002). Dr Jekyll and Dr Hyde: author-reader asymmetries in scholarly publishing. *Aslib Proceedings: new information perspectives*, Volume 54, Number 3, August 2002, pp. 149-157(9), Emerald Group Publishing Limited.
[37] Department of Trade and Industry: Publishing Sector: www.dti.gov.uk/sectors/publishing/index.html
[38] ALPSP & CAPP (2002). *Learned Journal Publishing in the UK. Briefing Prepared for UK Government*.

Indicative figures might also be usefully derived from analysis of both Ulrich's and Thomson Scientific's Journal Citation Reports (as a sub-sample) however, this is a significant research effort and care would need to be taken to assign appropriate imprints and subsidiaries to their parent publishers accurately.

Journal publishing revenues

There are three relatively widely cited sources that estimate the value of the STM publishing market in terms of publisher revenues:

- EPS Market Monitor, Scientific, technical & medical (STM) information: market trends and industry performance, June 2006 (Source 1.6);
- Outsell, I-Market MarketView: Scientific, technical & medical segment 2005, September 2005 (Source 1.10);
- Simba Information, Global STM market analysis & forecast 2004 (Source 1.11).

Each of these is global in focus and produces broadly comparable estimates of the overall value of the market. However, although they are similar in scope and thus provide opportunities for some cross-validation, their coverage of the journals publishing market does not map particularly well with the terms of reference of the current study. For example, each of these reports defines the market far more narrowly in discipline terms than the scope of this study (i.e. they each cover mainly or solely scientific, technical and medical content). Furthermore, in terms of geographic breakdown, there is some analysis at European level, but there is no estimate of publisher sales into the UK. Also, all of these sources encompass all formats of STM publishing (e.g. books, journals, aggregated databases) and hence journals tend to be dealt with only at a relatively cursory level of analysis. None of these three sources provides detailed revenue splits for journals (in terms of, say, customer breakdown or revenue stream for example).

A fourth source, the AAP/PSP Annual Industry Statistics Report, Section 4: Professional and Scholarly Journals, 2004 (Source 1.1) provides an alternative view of the supply-side marketplace. Unlike the three reports described above, this study presents primary data in the form of aggregated global sales data collected directly from a sample of 45 professional society and commercial journal publishers. The sample (listed in the report) is weighted towards the US, (particularly in terms of society publishers), but does include a number of the major global players to make its findings significant (data from Elsevier, Springer, Wolters Kluwer, John Wiley and Taylor and Francis are included). Whilst the survey sample is not weighted or structured to be statistically representative of the universe of global journal publishers, it includes enough of the major players to be relevant and useful. The combined journal publishing revenue of the participants in the survey is around $3.5 billion. This represents around 70 per cent of the mean of the EPS/Simba estimates for global STM English-language journals market (~$5 billion). Although the estimated value of the STM market is clearly only a sub-set – if a substantial one – of the total potential journals market (it does not take into account all disciplines or foreign language publisher revenues), the comparison suggests that the AAP/PSP data cover enough of the market to be valuable. Although the AAP/PSP report does not offer a 'whole' market estimate, such as that postulated by the three reports cited above, it offers deeper analysis in terms of journal output (journal titles, extent, articles, advertising pages), sales by type (e.g. paid circulation, articles, reprints, author fees) and unit sales by type (circulation, licences, single articles/pay-per-view). Some of these analyses can be usefully cross-validated with those of the other

three studies at the top level, but, unfortunately, for many of the deeper journal-specific data points, we have no data against which to test its likely accuracy.

Analysis/data validation

The analysis below is concerned both to validate and to triangulate key data points gleaned from the evidence base described above in Area 1. Where possible, we have either cross-compared data sources in order to validate robustness or sought to triangulate available data-points with other sources to fill key gaps or provide proxies or benchmarks. Clearly, this has not been possible on issues where only one reliable source has been identified. Some limited primary research has been undertaken to facilitate this analysis. When the research effort required is substantial and falls outside the remit of the present study, we have indicated what useful research might be undertaken.

Industry volume and value

Title and article output and trends

As discussed above, Ulrich's Periodicals Directory (Source 1.4) is the most cited source on journal title output. Title count data is presented in Mabe & Amin (Source 1.9) (1996) and Mabe[39] (2001). However, this does not provide us with a growth trend line as the data in Ulrich's database is not consistent over time. The editorial team add in titles that come on to their radar retrospectively, re-classify data and correct classification errors on an ongoing basis. Thus, if one runs a search on title counts for journals launched up to and including 1996 again today (2006), the search will produce a rather different result to that obtained by Mabe & Amin in their earlier research. Whilst Ulrich's is the most comprehensive and reliable source available to estimate the universe of journal titles, it must be used advisedly. With this in mind EPS[40] has checked the primary source to obtain an updated 2006 data point (using the search terms: scholarly journal + active + peer reviewed). This search produces a current count of 22,768 global journal titles – and this is the base data we have used in other analyses in this report.

Data extracted from Thomson Scientific's Journal Citation Reports (Source 1.13) suggest that around 1 million articles are produced annually by this core set of 8,700 scholarly journals. Internal research at Elsevier,[41] extrapolating this figure to the wider journals universe of circa 22,000 titles sourced from Ulrich's Directory, suggests an estimated total article output of 1.5 million. However, we have no way to validate this estimate.

Mabe & Amin (Source 1.9) have undertaken systematic secondary analysis of Ulrich's over time in terms of measuring journal title growth. They have established a compound annual growth rate of 3-5 per cent through to the mid-nineties.

Table 1.1: Growth in number of scholarly journals, 1900-1996

Period	Source	Author	Title compound annual growth rate (%)
1900-1944	Ulrich's 1999	Mabe & Amin	3.3%
1944-1978	"	"	4.7%
1978-1996	"	"	3.3%

[39] Mabe, M (2003). The Growth and Number of Journals, *Serials* Vol. 16, No. 2, July 2003.
[40] EPS primary research: www.ulrichsweb.com on 22/06/06
[41] Telephone interview with Mayur Amin, Director, Research and Academic Relations, Elsevier (Oxford) on 28/07/06

Mabe[42] re-visited and updated these growth analyses using the 2001 edition of Ulrich's and measured slightly different time periods through to 2000. However, clearly the overall underlying data set is similar and the CAGR outcomes are very much in agreement with the earlier research, at between 3-4 per cent.

Cox (Source 1.3) presents data quantifying journal launches (1,048) and closures/mergers (261) over the period 2001-2005 from a sample of 174 publishers. Clearly these data suggest a net increase in journals over the period. However, as the actual base number of titles published by these respondents was only collected in bands, it is not possible to cross-compare the actual five-year growth that this represents with Mabe's earlier trending data. The AAP/PSP (Source 1.1) sample of 45 primary publishers show a net 3 per cent gain in title output between 2003 and 2004, with output rising from 5,883 to 6,047, but clearly this represents only one year of output and cannot be seen as meaningful in terms of growth trend.

A further important sub-segmentation of the scholarly journal market is the number of titles which are online. Cox's (Source 1.3) survey data does not correlate particularly well with data drawn from Ulrich's (Source 1.4). It seems likely, however, that the geographical location of the publishers covered in the data accounts for this disparity, i.e. we can conclude that Ulrich's' figure is broadly representative of the global picture, while Cox describes a more advanced stage of online development representative of the UK and US publishing markets.

Table 1.2: Proportion of scholarly journals published online

Year	Source	Sample size	Sample location	All journals	STM	Arts & Humanities
2003	Cox	149 publishers	Circa 80% UK/USA 'location of publisher'	70%	83%	73%
2005	Cox	174 publishers	As above	90%	93%	84%
2006	Ulrich's	22,768 journals	Circa 50% UK/US 'country of publication'[43]	62%	-	-

Ulrich's (Source 1.4) and the Directory of Open Access Journals (DOAJ) (Source 1.8) can be cross-compared to quantify the volume of open access journals as a proportion of the total journals market. Further primary research would be required to measure growth over time. It should be noted that research by Morris[44] indicates that up to 14 per cent of DOAJ-listed titles may be in some way inactive. Cross-comparing this data with a search on Ulrichs database (for *active*, open access journals), the two sources seem broadly in line and suggest that there are around 2,000 active journal titles published under some form of open access model (and we should also note that some of those titles included in these totals may also be hybrid subscription/open access titles). This total represents around 9-10 per cent of peer reviewed scholarly journal title output as measured by Ulrich's in 2006.

Table 1.3: Number of scholarly journals published using open access model

[42] Mabe, M (2003). The Growth and Number of Journals. *Serials,* Vol. 16, No. 2, July 2003.
[43] EPS primary research: www.Ulrich'sweb.com on 22/06/06 (UK: 4,144 journals; USA: 7,746 journals)
[44] Morris, S (2006). When is a journal not a journal? A closer look at the DOAJ. *Learned Publishing*, Vol 19 (1), January 2006.

Year	Source	Sample overview	Number of refereed OA titles	All OA titles
2006	Ulrich's[45]	Sub-segment of base global sample of 22,768 *active*, peer-reviewed, scholarly journals	1,388	2,022
2006	Directory of Open Access Journals[46]	Directory of open access scientific and scholarly journals – all titles included are quality controlled (through an editor, editorial board and/or a peer-review system), but are not *necessarily* peer-reviewed.	-	2,309

Value of demand-side (i.e. spend by journal purchasers)

Spend by the UK-demand-side purchasing community is not quantified in the data in any coherent way; indeed, key elements of the evidence base are missing. As noted above, the expenditure statistics from SCONUL (Source 1.12) and LISU (Source 1.7) provide a breakdown of serials spending trends over a 10-year period, although the broader 'serials' category it measures will also include titles which are not scholarly peer-reviewed journals, such as professional and trade magazines. There are no comparable demand-side data for other types of libraries, or for individual journal purchasers at UK or global level. The AAP/PSP (Source 1.1) primary survey of 45 publishers (mostly US-based) estimates that sales to individuals (defined as subscriptions purchased for personal use) accounts for around 9 per cent of total sales revenues in 2004 (up 2 per cent from 2003). However, no other up-to-date trending data quantifying researchers' individual spend on journal content has been identified against which to validate this estimate.

According to LISU (Source 1.7), serial subscriptions accounted for 53 per cent (£94.5 million) of total UK library information service spend in the higher education segment in 2003-2004. Expenditure on scholarly peer-reviewed journals will be a sub-set of this spend, but is not broken out in the data. Spend on printed serials is declining as a proportion of total spend and now accounts for around 37 per cent of serials spend, while electronic-only journals account for 26 per cent and joint electronic/print subscriptions for 37 per cent of spend. The number of current serials subscriptions has more than doubled in HEI libraries in the last ten years to a mean value of 6,900 titles per institution (683 per 1,000 FTE students). LISU attributes much of this growth in title acquisition in recent years to 'bundled' deals, whereby libraries subscribe to packages of electronic journal titles from publishers at lower costs than the actual combined subscription prices. The extent and nature of this 'bundled' purchasing is not explored in detail in the data; LISU notes, however, that such deals are also partly responsible for lowering the average cost per title of current serial subscriptions by 23 per cent over the five-year period to 2003-2004. Data from the Association of Research Libraries in the US[47] shows similar trends with serial unit costs falling 15% since 2000 in tandem with an increase over the same period in the total number of serials purchased by member libraries (bundled deals, consortial arrangements and the increasing availability of open access titles may all have contributed to this trend, but no clear evidence is provided).

If we take the supply-side customer segmentation from both the EPS STM Market Monitor (Source 1.6) and the Simba Global STM market analysis (Source 1.11)

[45] CSA, *Ulrich's Periodicals Directory*.
[46] *Directory of Open Access Journals*: www.doaj.org as at 17/07/06
47 ARL, Monograph and Serial Expenditures in ARL libraries, 1986-2004: http://www.arl.org/stats/arlstat/

and extrapolate the expenditure data from the UK HEI library segment, we can postulate very approximate ranges of total UK market spend. It is important to bear in mind that the £94.5m figure estimated by LISU for UK HEI serials expenditure (2003-2004) includes a broader range of journal titles than our definition (peer-reviewed, scholarly journals). Hence, any further calculations derived from this base number is likely to overestimate the total market. Clearly this indicates a key gap in the evidence base in terms of actual data (i.e. spending levels of journal customer groups beyond the HEI segment). The resulting estimates of total UK market expenditure range between £192 million ($348m) and £292 million ($526m); they vary widely because of differing scope of the base reports and the differing estimates of the proportionate spend accounted for by the academic segment. UK market spend on journals in these ranges would represent approximately 7-11 per cent of global journal market sales (according to total market estimates described below; 'value of supply side').

Table 1.4: Extrapolated data to estimate a potential range of UK scholarly journals market spend (£)

	STM Revenue breakdown (%) by customer type EPS (2004)	Demand-side spend (extrapolated)
Academic	49.0%	**£94.5m**
Corporate	31.0%	*= £59.6m*
Government	5.0%	*= £9.6m*
Professionals/practitioners (medical only)	15.0%	*= £28.8m*
Total		*£192.3m*
	STM Revenue breakdown (%) by customer type Simba (2004 *Projection*)	
Academic	32.3%	**£94.5m**
Corporate	19.4%	*= £56.6m*
Individual	13.2%	*= £38.6m*
Government	6.7%	*= £19.6m*
Society members	4.7%	*= £13.8m*
Hospital	2.6%	*= £7.6m*
Advertising/promotions/others	21.0%	*= £61.4m*
Total		*£292.1m*

[Italics: extrapolated values based on UK HEI spend in 2003-2004 of £94.5 m: does not total exactly]

Acquisition and cancellation trends

As reported above, there are few published quantitative data in this area and nothing that can be usefully cross-validated. In the area of journal pricing there are, however, a number of databases which do merit consideration.

Journal pricing

Van Orsdel & Born (Source 1.15) and LISU (Source 1.7) provide us with top-level data on pricing trends using average prices derived from large journal title samples to track increases over time (along with variance by geography and discipline). Unfortunately, the data are split rather differently and this, along with potential currency differences and differences in overseas versus domestic prices (overseas prices may include additional carriage costs for print), makes it hard to cross-validate the findings. At the very top level, however, price increase data for 04-05 are broadly in line – somewhere between 7 and 10 per cent across the board. Cross-analysis of longer term trends would potentially be more indicative of consensus, but this would be a more substantial research exercise outside of the remit of the current study.

Disciplinary analyses from the two sources seem broadly to correlate as regards relative pricing between the different disciplines, positioning arts and humanities titles as substantially cheaper than titles in the sciences, and the hard sciences titles as the most expensive of all the categories. There is a marked price differential between US and non-US publishers in the Van Orsdel & Born data which is likely explained to some extent by either currency adjustments or 'overseas' pricing differentials, but no supporting data or explanations on this issue are provided.

*Table 1.5: Average journal prices by discipline, 2005**

Source	Sample	Location of publisher	Social Sciences	Arts & Humanities	Science	Technology	Medicine
LISU/Swets	Swets database of >67,000 periodicals	GB	£494 $889	£98 $176 (humanities only)	£780 $1,404	£414 $745	£514 $925
					STM		
Van Orsdel & Born	Sub-sample from Thomson Scientific Journal Citation Indexes: 4,893 titles included	USA	$349	$162	$1,068		
As above	As above	Non-USA	$721	$235	$1,732		

** Pound Sterling to US Dollar conversion rate of £1=$1.8 has been applied*

Value of supply side (i.e. publisher revenues)

The three main sources of data on supply-side revenues in the scholarly journals market from EPS (Source 1.6), Simba (Source 1.11) and Outsell (Source 1.10) each provide a number of data segmentations. In the analyses below we attempt to cross-compare their estimates. As noted above, none of the sources identified provides data relating specifically to the market for peer-reviewed, scholarly journals across all disciplines (the focus of the current study). Each source takes the narrower disciplinary perspective of scientific, technical and medical

publications only, but in product terms, covers the market much more broadly including books and database products in their analyses.

Table 1.6: Summary of source coverage (supply side analysis)

Simba (2003)	MarketMonitor (2005)	Outsell (2004)
Global revenues for English-language products in the scientific, technical and medical market. Includes harder-edge social sciences (e.g. economics). Includes: books, journals, abstracting and indexing databases.	Global revenues of European and North American publishers of scientific, technical and medical content. Includes: books, journals, secondary information services and syndication of non-textual scientific and technical data.	Global revenues of publishers (not defined) of scientific, technical and medical content. Includes: journals, books, other technical content (e.g. patents, standards) and secondary aggregation services.

Estimates of the overall global STM market are broadly in line. The lower estimate calculated by the Market Monitor is a result of its narrower definition of the marketplace.

Table 1.7: Global revenues generated by publishers in STM disciplines (all products, STM only)

	2003 ($ billion)	2004 ($ billion)	2005 ($ billion)
Simba[48]	$11.5	$12.0	-
Market Monitor	$8.7	$9.3	$9.7
Outsell	$11.5	$12.4	-

Estimates for global sales in the scholarly journal segment are similarly comparable, both Simba and Market Monitor estimate journal sales to account for approximately 45-50 per cent of the total STM publishing market and provide a 2004 market value of $4.7-5.4 billion. Year-on-year revenue growth is estimated at between 6 per cent (Simba 2003-2004) and 12 per cent (Market Monitor, 2003-2004).

Table 1.8: Global revenues generated by scholarly journals in STM disciplines

	2003 ($ billion)	2004 ($ billion)	2005 ($ billion)
Simba (total STM)	**$5.1**	**$5.4**	-
Science & technology	*$2.7*	*$2.9*	-
Medicine	*$2.4*	*$2.5*	-
MarketMonitor (total STM)	**$4.2**	**$4.7**	**$4.9**
Science & technology	*$2.4*	*$2.7*	-
Medicine	*$1.8*	*$2.0*	-
Outsell	No sub-analysis of journals market included		

Split of publisher revenue between customer segments

None of the three key sources provides customer segmentation data specifically for the journals market. At the aggregated STM market level, it is more difficult to cross-compare the data, as the report authors have segmented the market slightly differently according to their particular market definitions and scope. Our inability to split these data specifically by journals and, within that, by disciplines means that some characteristics of the journals market are masked. For example, advertising, promotions and sponsorship are only relevant to a small sub-set of journals, but for these players can be a very significant revenue stream. Similarly, the corporate market is probably more important as a

[48] Simba: 2004 data are projected.

purchaser of content in some areas than others for example, it is probably insignificant in revenue terms for social sciences and humanities journal sales.

Table 1.9: Global STM revenues by customer type

	Simba (2004P)	MarketMonitor (2004)	Outsell (2004)
Academic	32.3	49.0	31.7
Corporate	19.4	31.0	36.1
Individual	13.2	-	-
Government	6.7	5.0	5.0
Society members	4.7	-	-
Hospital	2.6	-	-
Advertising/promotions/others	21.0	-	-
Professionals/practitioners (medical only)	-	15.0	-
Healthcare	-	-	27.2

There is a more uniform approach to sizing sales by geographic region. Again, however, no journal level sales break-downs are included in the data nor is any specifically UK-focused split provided. Estimates for sales into the Europe/EMEA region range between 24 and 32 per cent of the total STM publishing market.

Table 1.10: Global STM revenues by geography

	Simba (2004P)	MarketMonitor (2004)	Outsell (2004)
USA/North America	63.6	53.0	54.1
Europe	24.2	31.0	-
Asia Pacific	10.2	9.0	8.2
Rest of World	1.9	6.0	5.7
EMEA	-	-	32.0

Using an average of the journals market estimates cited above (~$5 billion), this suggests European STM scholarly journal sales of between $1.2 billion and $1.6 billion in 2004, although this assumes that the geographical breakdown for journal sales is the same as for the full range of STM products.

Split between print and electronic revenues

Data recording the split between print and electronic revenues in the scholarly journals field is almost impossible to calculate accurately in the current marketplace due to diverse models for 'bundling' print and electronic pricing and lack of consistency in terms of publisher accounting practices when allocating revenues for these types of packaged products.

Despite these challenges, various sources have sought to provide some kind of benchmark on this issue. EPS and Outsell each provide an estimated breakdown of sales of the total STM publishing market by format. The EPS and Outsell data provide a relatively consistent picture of the revenue split for the overall market, but, again, the scope is too narrow in disciplinary terms and too broad in product terms to accurately represent the scholarly journals market.

Table 1.11: Total STM publishing market: revenue split by format (2004)

	Electronic	Print	Other
EPS (Source 1.6)	56%	44%	-
Outsell (Source 1.10)	49%	48%	3%

The AAP/PSP (Source 1.1) and Waltham[49] (Area 2; Source 2.11) reports are more relevant providing a break down journal sales in terms of delivery format, based on direct publisher input. Furthermore, the primary survey sources relate specifically to sales of journal subscriptions (other revenue sources such as single-article sales and advertising sales are excluded).

Table 1.12: Journal publishing market: subscription revenue split by format (2004)

	Sample	Electronic	Print	Print and electronic
AAP/PSP	45 publishers; learned societies and commercial, US-based and international, multiple disciplines (but not structured sample), 6,047 journals covered	22%	21%	57%
Waltham	7 learned society publishers; primarily UK-based, STM disciplines only, 9 journals covered	1%	13%	86%

However, the results from these two surveys do not match particularly well. Clearly the AAP/PSP data are from a significantly broader sample in terms of both type of publishers and size of publisher and is thus likely to reflect the true market situation more closely. Combined print and electronic subscriptions dominate in both cases, but the shift towards electronic-only subscriptions is far more marked in terms of sales revenues in the larger sample. Unfortunately, we do not have a complete enough picture of demand-side data to consider alongside these supply-side splits. For the UK HEI segment, we reported above that spend splits out as: printed serials 37 per cent, electronic-only journals 26 per cent and joint electronic/print subscriptions 37 per cent (LISU – Source 1.7). These expenditure figures clearly show a greater weighting towards print spend than the AAP/PSP revenue splits. However, this may be due to geographic weighting of the samples (VAT is added to electronic title sales in the UK and this affects purchasing behaviour), the much broader definition of serials (as opposed to scholarly journals) in the UK HEI data, and the influence of other customer segments (healthcare, corporate and government) not accounted for at all in the demand-side estimates. Unfortunately ARL's data on the US HEI research library market estimates electronic serials only as a proportion of total spend on serials (41 per cent), so this does not offer an alternative source against which to map the AAP/PSP estimates.

[49] Waltham, M, on behalf of JISC (June 2005). Learned Society Open Access Business Models. Note: Waltham's survey of learned society publishers collects detailed data about the operation of a sample of thirteen journals from nine publishers. However, the number of respondents contributing valid data varies according to the particular question asked. Thus, throughout this study we have indicated the base number of respondents included in each case.

Value of PPV article sales

Two sources contribute data on this issue. Firstly Waltham (Area 2, Source 2.11) reports on the breakdown of revenues by source for 10 learned society journals. These data shown in Table 1.15 below, reveal that subscription revenue is by far the dominant source of income for these journals. Pay-per-view income is subsumed into 'other sources' and therefore accounts for less than 6 per cent of journal revenue for this sample of journals in 2004.

Table 1.13: Breakdown of publisher sales by revenue source: Waltham (2002-2004)

Year	Revenue from subscribers as % of total revenue	Revenue from authors as a % of total revenues	Revenue from other sources as % of total revenue
2002	89	4	7*
2003	89	5	6*
2004	88	6	6*

**Includes pay-per-view income.*
[Source: Waltham, 10 journals covered in data]

Data from the AAP/PSP primary survey of 45 publishers (Source 1.1) measures single article sales in 2003-2004 at 2.9 per cent of total income and correlates reasonably on this point with the Waltham data. This suggests that single article sales accounted for approximately $145 million of total journal publisher revenues of ~$5 billion in 2004. According to the AAP/PSP data, unit sales of single articles are growing, with an increase of over 40 per cent in single electronic article downloads between 2003 and 2004. Results from the AAP/PSP sample shows slightly less reliance on subscription income and a wider range revenue streams.

Table 1.14: Breakdown of publisher sales by revenue source: AAP/PSP (2004)

Year	Revenue from subscribers as % of total revenue	Revenue from authors as a % of total revenues	Revenue from single article sales	Revenue from advertising	Revenue from reprints
2004	77	1	3	12	7

*[Source: AAP/PSP, 6,047 journals covered in data. *Total journal publishing revenues of sample in survey reported as $3.5 billion]*

Split between commercial and society publishers (volume and value)

As reported above, the literature trawl has not revealed a reliable breakdown of publisher numbers by type, but a close analysis of the membership of journal publishing-related trade associations internationally would be informative. ALPSP has already undertaken some ground-clearing research in this area internally, but this is, as yet, unpublished.

There is little evidence on journal title output and segmentation by publisher type quoted in the literature. In their response to the Science & Technology Committee Inquiry into Scientific Publications in 2004, ALPSP quotes a market estimate of around 9,250 journals published by non-profit distributing organisations (such as a learned society, professional association, charitable or government institution or university presses). This estimate[50] is a helpful indicator, but must be treated with caution; it is a subset of the Ulrich's database

[50] Search of UlrichsWeb undertaken by the British Library in 2003 filtering within the search term 'publisher' for the terms 'society', 'association', 'institute' and 'university'; no foreign or variant spellings included.

of active, peer-reviewed scholarly journals, but (a) may not have captured all relevant journals, (b) may include some titles published on behalf of a non-profit distributing organisation by a commercial publisher and (c) will not have captured organisation names listed in the database in languages other than English. Some unpublished sources also merit consideration, but cannot be formally evaluated for this report. In terms of title output, ALPSP[51] undertook an analysis of the 500 most cited journals in the Thomson Scientific Journal Citation Reports in 1999 and found that 69% of these were published by non-profit distributing organisations. Clearly, this kind of analysis could be relatively easily updated.

Detailed bespoke analysis of relative article output of the different players has been undertaken internally at Elsevier[52] based on data derived from Thomson Scientific's Journal Citation Reports (Source 1.13). These data break down article output for this core sub-set (circa 8,700) of scholarly journals between commercial publishers (this category includes contract publishing on behalf of learned societies), university presses and learned societies.

Table 1.15: Percentage of total journal article output by type of publisher

Type of publisher	% of article output
Commercial publishers (incl publishing for societies)	64%
Society publishers	30%
University presses	4%
Other publishers	2%

[Source: Elsevier bespoke analysis of Thomson Scientific data]

Some data on publisher revenues do exist. Market Monitor (Source 1.6) looks at the total STM publishing revenue base of around 130 of the largest global learned society publishers. This list is dominated by US players, the largest of which are the American Chemical Society, the American Medical Association and the American Institute of Physics. It is estimated that the five largest non-profit distributing society publishers generate around 6 per cent of total STM market revenues. The remaining 125 large players included in the monitor are estimated to account for a further 8 per cent, Market Monitor suggests that society publishers generate approximately 14 per cent of total STM publishing revenues.

Gaps in the data

A key objective of this study is to illuminate gaps in the primary evidence base relating to the size and structure of the scholarly journals market, as well as to highlight:

- where such gaps can be filled through further research; and
- where there are issues which can not be proved for certain either way.

The review of sources available to contribute evidence to the issues surrounding journal demand-side economics and market size clearly illustrates a lack of fundamental data relating to numbers of journals and publishers. As even data at the very top level are informed 'guesstimates' it is clear that we have no firm grasp on important deeper segmentations of the market in terms of both journal and publisher diversity. The general nature of reporting regarding the size and structure of the journals publishing market hinders our interpretation of important trends in this marketplace. Thus, analyses regarding sales to different types of

[51] ALPSP, unpublished analysis of ISI (now Thomson Scientific) Top 500 Cited Journals undertaken in 1999.
[52] Telephone interview with Mayur Amin, Director, Research and Academic Relations, Elsevier (Oxford) on 28/07/06.

customers, sales in different formats, different types of subscriptions, or growth in individual article sales, for example, can currently be understood only at a very high level with no understanding of the trends in the different market segmentations which drive these data.

The main sources of supply-side data are, at best, indicative in the context of this study, as they do not report directly on the peer-reviewed scholarly journals market. The AAP/PSP (Source 1.1) data are highly valuable in terms of the depth they are able to offer by virtue of being directly sourced from the publishers themselves and it is clear that survey work of this type – extended to include as balanced and comprehensive a sample of the publisher universe as possible – is much needed, and would go a long way to filling the gaps in the evidence base on market size and structure.

Sources evaluated in detail in Appendix 1

Source 1: AAP/PSP Annual Industry Statistics Report: Section 4: Professional and scholarly journals, 2004 (data only, not full report)

Source 2: Bergstrom, T & McAfee, P. Journal Cost Effectiveness Web Site: http://www.journalprices.com/

Source 3: Cox J and Cox L (2006). Scholarly Publishing Practice – The ALPSP report on academic journal publishers' policies and practices in online publishing (Second Survey)

Source 4: CSA, Ulrich's Periodicals Directory (Ulrich'sweb: www.Ulrich'sweb.com)

Source 5: Davis, Philip and the Cornell University Library Task Force on Open Access Publishing. Dec 22, 2004. Calculating the Cost per Article in the Current Subscription Model: http://people.cornell.edu/pages/pmd8/

Source 6: EPS Market Monitor (2006). Scientific, technical & medical (STM) information: market trends and industry performance, June 2006

Source 7: LISU Annual Library Statistics 2005, Claire Creaser, Sally Maynard and Sonia White, November 2005

Source 8: Lund University Libraries, The Directory of Open Access Journals: http://www.doaj.org

Source 9: Mabe M & Amin M, Elsevier Science, Oxford, UK, (2001). Growth Dynamics of Scholarly and Scientific Journals. Scientometrics 51 (1) 2001

Source 10: Outsell, Inc. (2005). I-Market MarketView: Scientific, technical & medical segment 2005, September 2005.

Source 11: Simba Information (2004). Global STM market analysis & forecast 2004

Source 12: Society of College, National and University Libraries (SCONUL) (2005), Annual Library Statistics 2003-2004

Source 13: Thomson Scientific (ISI). Journal Citation Reports and Indexes.

Source 14: University of Wisconsin, Madison Campus Libraries. Journal Value Project (JVP), http://www.wendt.wisc.edu/projects/jvp/welcome.do

Source 15: Van Orsdel, L & Born, K, (2005). Periodical Price Survey. Library Magazine, April 15, 2005

Source 16: Ware, M, on behalf of ALPSP (March 2006). Survey of librarians on factors in journal cancellation

Area 2: Journal supply-side economics

Executive summary

Key questions

The key issues under examination in this area of the report are:

1. Evidence about the effort and cost incurred on the one hand by publishers and on the other by academics as part of their respective contributions to the publications process (for instance, costs associated with editorial and quality issues – including both the external and internal aspects of peer review).
2. Data that might help in developing understanding of any differences that there may be in journal economics between commercial and learned society publishers.
3. Reliable data on the costs of launching new products (both direct and indirect) and on the factors determining the investment required for new launches.

The evidence

Estimates suggest that 'content creation' costs, sometimes called 'first copy costs' (which do not vary with either the volume of output or the number of subscribers served) vary widely from publisher to publisher with data points collected ranging from $250 to $2,000 per published article. Baseline data relating to specific parts of the process are rare, including specific costs relating to peer review.

In addition to these 'first copy article costs', publishers incur both variable costs relating to production and distribution and general fixed overheads. Variable costs for journal publishing are very difficult to approximate as they vary significantly according to the nature of the content (illustrations, colour, graphics can all impact on costs), journal extent and frequency, distribution medium and circulation.

The variation in the data points collected in the existing evidence base (see chart below) indicates that a 'broad-brush' approach to the cost profile of scholarly journal publishing probably oversimplifies the picture too much to be useful.

Journal publishing cost element	Data ranges
Content creation costs	16%-54%
Overheads	11%-55%
Manufacturing, printing and paper	8%-40%
Distribution and fulfilment	3%-17%

The gaps

We have some useful base data on costs of the current system of journal publishing. There are gaps and weaknesses, but, in the area of supply-side economics, many of these are in fact capable of being overcome through further research.

Very little up-to-date evidence is available regarding the cost (in terms of time or money) of the researcher community's contribution as authors, editors or peer reviewers to the current scholarly journals publishing process.

Evidence relating to the costs of launching new journals is also lacking.

The evidence base relating to publisher costs is partial in terms of its coverage of the publishing community (commercial and university presses in particular are under-represented) and it is not segmented sufficiently in terms of processes to address the particular questions being reviewed in this area of the current study.

The evidence of diversity in journal publishing economics provided by existing sources suggests that this issue will be most productively investigated at the journal level, and that data giving ranges of costs by discipline, frequency, extent and circulation will most accurately reflect the true complexity of supply-side costs.

Introduction

This review of journal supply-side economics focuses on examining evidence relating to the costs of the existing publishing process. Key questions identified for investigation include understanding the various roles (and costs) of both academic and publisher participation in this process and the potential differences between costs incurred by non-profit distributing publishers and commercial publishers. Primary research that identifies and quantifies the costs specifically related to the launch of new journals has also been examined.

In summary, the issues specified for coverage in Area 2 of the report are:

4. Evidence about the effort and cost incurred on the one hand by publishers and on the other by academics as part of their respective contributions to the publications process (for instance, costs associated with editorial and quality issues – including both the external and internal aspects of peer review).
5. Data that might help in developing understanding of any differences that there may be in journal economics between commercial and learned society (or other non-profit distributing) publishers.
6. Reliable data on the costs of launching new products (both direct and indirect) and on the factors determining the investment required for new journal launches.

Scope and definitions

Journal supply-side economics is defined in this report as the costs and effort incurred through the process of creating, producing and distributing scholarly journals to end-users (i.e. the research community). We have reviewed the existing primary research that examines the relative contribution and costs of the three major stakeholders in this process:

- Researchers (in both academic and corporate environments);
- Publishers (both non-profit distributing and commercial); and
- Librarians (in both academic and corporate environments).

Under the umbrella heading of 'journal publishing supply-side economics' we are considering the costs of the different steps involved in creating and distributing journal content. This area is defined in this study as including consideration of data relating to providing access to publisher back-files. We do not, however, include the issue of long-term preservation (i.e. the archiving function) in our costs review (although it is noted that commitments with regards to the legal deposit of electronic content may ultimately mean increased costs for publishers, if it requires submission of content in a standardised format for example). Long-term preservation, typically undertaken by national, deposit and other academic libraries in the UK, is integral to the scholarly communications process in its widest sense. However, it is a complex area in its own right and is deemed outside of the scope of this study.

Key sources

Costs of academic, publisher and library participation

Primary research relating to journal supply-side economics is limited. Most authors who consider the issue cite the reluctance of publishers to release cost data as the main explanation for the paucity of real evidence. In fact, the published literature contains only two recent analyses of cost data collected from publishers. Furthermore, the sample sizes responding to each are very limited.

Dryburgh[53] (Source 2.2) examines the cost base of ten book and journal publishing operations (seven of these are learned societies, two are university presses and one is a commercial publisher). These ten survey respondents are named in the report, but the data themselves are presented anonymously. The study was conducted on behalf of the publishing association ALPSP and sought to collect financial indicators in a consistent way in order to present a set of benchmarks of financial efficiency for cross-comparison and use by other publishers. The sample is very small and is skewed in terms of disciplinary focus (nine of the ten respondents publish in science); all respondents are UK-based publishers. The participants do, however, represent a reasonable range of organisations in terms of size – numbers of titles published ranging from one to 140 and numbers of papers published from 100 to 12,000. Clearly, however, none of the larger players is represented here. The data collected provide insight into the direct costs of journal publishing and provide cross-comparisons of per page and per article costs for different publishers. Variable costs and overheads related to journal publishing are less well covered in the survey.

Waltham[54] (Source 2.10) presents a cost analysis for a sample of thirteen learned society journals from nine publishers. This study was commissioned by JISC to find out if, and how, learned society publishers can consider making a transition to a sustainable open access business model. The focus of the report is an in-depth exploration of the business and pricing models of thirteen journals. This self-selected sample of publishers consists of eight UK-based societies and one US-based publisher all of whom publish either in science or technology. Participants provided financial data to an agreed template including detailed journal-level cost and revenue data. Some additional analyses have been undertaken on the reported data for the purposes of the current study. It is important to note that both the number of respondents and journals contributing

[53] Dryburgh, A (2002). The costs of learned journal and book publishing: a benchmarking study for ALPSP.
[54] Waltham, M, on behalf of JISC (June 2005). Learned Society Open Access Business Models.

valid data varies according to the particular question asked. We have quoted the base number of respondents and journals included for each data point reviewed.

We would expect a reasonable degree of homogeneity in publishing costs across similarly sized organisations. Thus, as both of these survey samples include a reasonable range of different-sized organisations, they are likely to provide an indicative and useful range of publisher cost estimates. However, the fact that non-profit distributing publishers account for over 90 per cent (18/19) of survey participants in these two studies combined is a serious limitation and points to a significant gap with regard to data relating to the cost profile of commercial journal publishers. Furthermore, the very largest publishers are not covered in either study. It should also be noted that publishing processes are not identical from organisation to organisation. Some publishers may chose not to copy-edit, some internally review manuscripts before forwarding them on for peer review while others do not. Thus, it should be appreciated that the chain of publishing processes is not always uniform; ideally, a cost review at journal level would provide costs broken down by each process to take account of this variability. However, the current evidence base does not permit this level of detailed analysis.

Primary analysis of publisher peer review costs from Donovan[55] was disregarded owing to the lack of transparency regarding method and scope (it is unclear whether organisations surveyed include or exclude overheads in their cost estimates and the data can not be cross-compared). Secondary analysis of both Donovan and of other studies by Rowland[56] has very little robust data on which to draw and is likewise not included in our analysis.

Further sources provide insight into journal publishing costs through secondary analyses of available primary data. Tenopir and King[57] (Sources 2.6, 2.7, 2.8, 2.9) have used many years of primary research of the US scientist and publisher community to build a model of cost estimates for journal publishing. SQW Consultants[58] (Source 6.7), on behalf of the Wellcome Trust, seek to assess the actual costs of publishing scientific, technical and medical research in peer-reviewed journals and to compare the costs of the current 'subscriber-pays' model to costs for an 'author-pays' open access model, drawing primarily on available primary data (the authors draw heavily on Dryburgh and Tenopir & King) and discussions with publishers.

The SQW report provides a useful overview of the existing literature on 'first copy' journal publishing costs, but does not address the question of overhead and other costs. Morris[59] (Source 2.3) considers the full costs of research communication under the existing subscription model, taking into account costs for the research itself, writing the journal paper, publishing the paper, acquiring the paper, reading the paper and preserving the paper. The article draws together previous research on the costs of these different activities and updates them to 2004.

[55] Donovan, B (1998). The Truth About Peer Review, Learned Publishing, 11(3), 170-184.
[56] Rowland, F (2002). Report for the JISC Scholarly Communications Group. The Peer Review Process.
[57] Tenopir, C & King, D, (1998). Designing Electronic Journals with Thirty Years of Lessons from Print, The Journal of Electronic Publishing, December, 1998 Volume 4, Issue 2. Tenopir, C & King, D (1998). Economic Cost Models of Scientific Scholarly Journals, University of Tennessee School of Information Sciences Center for Information Studies Paper presented to the ICSU Workshop, Keble College, Oxford, UK, 31 March to 2 April 1998. King, D & Tenopir, C (2004). An Author Based Assessment of Author Pays, Nature Web Focus. Tenopir, C and Donald, K, (2000). Towards Electronic Journals: Realities for Scientists, Librarians, and Publishers. Washington, D.C.: Special Libraries Association.
[58] SQW, Commissioned by the Wellcome Trust (April 2004). Costs and Business Models in Scientific Research Publishing.
[59] Morris, Sally (2005). The True Costs of Scholarly Publishing, in Learned Publishing, 18(2), 115-126.

Proxies are used when detailed raw data are lacking and whilst this implies that the results are indicative rather than robust, the additional analyses add value to the raw data and provide context in which to cross-validate the primary sources. Mellman[60] adds an interesting case study of a single journal to the debate.

There has been very little primary research into the costs of the journal publishing system incurred by academics – whether as 1) authors, 2) reviewers, or 3) editors. Tenopir and King (Sources 2.6, 2.7, 2.8, 2.9) derived a range of metrics relating to cost estimates for US scientists' time spent writing, reviewing and reading from their research during the 1990s. Morris (Source 2.3) provides a useful summary and updated interpretation of this data, but no new primary research has been uncovered in the literature search.

We have also looked at the costs incurred by libraries in supplying content to end-users, i.e. the non-subscription related costs to libraries of the current journals system (e.g. cataloguing, licensing negotiations, subscription processing and so on). These data are valuable in terms of both tracking the impact on library costs of the shift to electronic formats and considering the potential impact of alternative business models. LISU[61] reports that, on average, libraries in the UK HEI sector spend around a third of their total expenditure on information, with the remainder being spent on staff and also, to a lesser extent, on equipment and other expenses. A recent report by Schonfeld *et al.*[62] (Source 2.5) contributes data on the non-subscription related costs of 11 US academic libraries. However, no comparable UK academic library data or data relating specifically to the corporate library environment have been identified.

Differences between economics of commercial, learned society, and university press journal publishers

The fact that there are no primary data available on journal publishing costs for commercial players or university presses precludes any detailed analysis of the key differences in economics between these different players. As noted above, we have useful data from Waltham (Source 2.10) relating to the costs, revenues and business models operated by the learned society segment of this market. Baldwin[63] (Source 2.1) adds to this evidence base with a further survey of professional and learned membership bodies, which examines how any surplus revenues generated by these publishing activities are spent. Analyses in this report also make the important distinction between society publishers who run their journals internally and those who subcontract core publishing functions to a commercial publisher.

Costs of launching new journals

Though now somewhat dated, Page *et al.*[64] (Source 2.4) is the most valuable source of data relating to the costs of launching new journals. With this exception, the literature provides only anecdotal evidence from individual

[60] Mellman I, (2006). How Journals Can Realistically Boost Access, Nature Web Focus, July 11th, 2006
[61] Creaser, C, Maynard, S, White, S (November 2005). LISU Annual Library Statistics 2005.
[62] Schonfeld, King, Okersen and Fenton (2004) on behalf of the Council on Library and Information Resources. The Non-Subscription Side of Periodicals: Changes in Library Operations and Costs between Print and Electronic Formats.
[63] Baldwin, C (2004). What do Societies do with their Surpluses? ALPSP/Blackwell Survey.
[64] Page, G, Campbell, R, Meadows, J (1997). Journal Publishing. Cambridge University Press, pp 407 ISBN 0 521 44137 4.

publishers. This is an area in which no detailed primary research has been undertaken.

The key sources drawn on in this section have all been evaluated to assess their applicability to the study. A list of these sources is included at the end of this section and individual source evaluation sheets for each one are included in Appendix 1 of this report.

Analysis/data validation

Costs of academic, publisher and library participation in the journal publishing process

Academics

Researchers and academics are reported to contribute to the publishing process in three main ways:

1. As authors, by writing and submitting article manuscripts;
2. As editors or editorial board members for a specific journal; and
3. As peer reviewers.

Limited data have been collected relating to the time and cost of researchers' participation in these activities. Tenopir and King (Source 2.8) surveyed large samples of US scientists between 1993 and 1998. They reported that only about 15-20 per cent of scientists in the United States have authored a refereed article. This early research found that these authors and co-authors were spending a total of around 80-100 hours preparing each manuscript. Based on scientist salary levels at the time, Tenopir and King estimated an average author cost of around $6,000 per article (Morris [Source 2.3] updates this figure to reflect inflation in 2004 to $6,700 per article). However, most of this work pre-dates the introduction of electronic editorial office management systems, online submission of manuscripts and electronic peer review. Cox[65] (see Area 1; Source 1.3) reported that over 50 per cent of 174 publishers in his survey are using online peer review systems and almost all larger publishers do so. The introduction of these new editorial practices and tools could potentially have impacted the time involved in manuscript preparation for researchers since Tenopir & King's early research. Research conducted by King in 2003 with 2,500 scientists in higher education at the University of Pittsburgh[66] reports an average of 95 hours per article. Whilse this sample is limited to US researchers only, it seems to suggest that the time spent in preparing manuscripts has in fact changed little over time. We do not have any UK-specific data against which to cross-validate these data, but there is no particular reason to suggest that time spent preparing manuscripts for submission varies internationally.

We have not discovered any primary data measuring the time invested by researchers in fulfilling an editorial role on behalf of journal publishers, or indeed any comprehensive evidence relating to the different ways in which they are recompensed for their contribution. It is reported in Morris (Source 2.3) that publishers usually cover the costs of office accommodation and other expenses for their scholarly journal editors and often pay an honorarium or royalty in addition (for example, editors of medical journals may be paid at the relevant

[65] Cox J and Cox L (2005). Scholarly Publishing Practice – The ALPSP report on academic journal publishers' policies and practices in online publishing (Second Survey)
[66] King, D & Tenopir, C (2004). An Evidence Based Assessment of Author Pays, Nature Web Focus.

National Health Service rate for one or two sessions per week). The journal *Nature* is quoted as being exceptional in that it does have academic editors on its payroll. However, these reports are based on anecdotal sources only and there is no primary research to substantiate or provide further detail on either the costs incurred or time spent by researchers in undertaking an editorial role *per se* or relating to the actual contribution to expenses made by publishers.

However some work has been done – again by Tenopir and King (Source 2.8), and quoted in Morris (Source 2.3) – estimating the costs of the peer review process to the individual researcher. Their research suggests that the input of 2-3 reviewers each spending an average of 3-6 hours reviewing a manuscript costs $480 (updated by Morris to $540 per paper adjusted to 2004 rates). This process is often iterative (the reviewer will raise queries directly with the author, who will respond to the reviewer to debate these issues and clarify certain points). Again no further primary data are available against which to cross-validate this calculation.

Publishers

Content creation ('first copy') costs

Two UK-based publisher surveys, Dryburgh (Source 2.2) and Waltham (Source 2.10) provide us with base data regarding journal publishing costs. Further work by Tenopir and King (Sources 2.6, 2.7, 2.8, 2.9) and additional secondary analysis by SQW (Source 6.7) and Morris (Source 2.3) provide useful benchmarks for cross comparison.

Research on journal publishing costs makes a distinction between the *fixed* and *variable* cost elements involved. The distinction is not straightforward however. For example, there are differences between print and electronic cost structures (for example, mailing costs for printed journals go up proportionately with subscriber volumes, but online distribution costs are hardly affected at all, so the variable cost impact is very different in the print and online worlds). However, the sources reviewed are broadly in line in terms of identifying and categorising the main cost elements as described below.

Table 2.1: Allocation of journal publishing costs

Publishing support and general overheads	Article and issue creation costs	Production and distribution
Journal publishing system support Appointing and managing editor/editorial board and then managing changes (when editors resign, retire, are changed) Management of other revenue streams (reprints, off-prints, author fees) Rights management (including legal permissions and contract management) Sales and marketing (including licence negotiations, promotion to authors, promotion to libraries) Develop, maintain and update online systems Provision of usage statistics Data conversion Managing the journal list: divesting and acquiring titles, contract negotiations between societies and publishers Launch new journals **General overheads** General management (e.g. HR, finance, strategy planning)	**Article creation costs ('first copy costs')** Review and manage submitted manuscripts Manage peer review Support authors Copy-edit/rewrite article content Edit/manage illustrations Quality assurance of e-content (and any multimedia content) and checking metadata Type-setting/page formatting **Per issue costs** Create and copy-edit non-article content Edit/manage illustrations Quality assurance of e-content (and any multimedia content) and checking metadata Type-setting/page formatting Issue compilation	**Printed journals:** *Production:* Printing Paper Binding *Distribution:* Fulfilment (mailing costs) Subscription management Customer service (including claims and global technical support) **Electronic journals:** *Production:* Upload to server and provide ongoing online hosting/storage *Distribution:* Subscription management Customer service (including claims and global technical support)

Article creation costs are often described by the journal publishing industry as 'first copy costs'. This refers to the costs incurred in getting a copy of a given article into the state required for it to be published in the journal or uploaded to a server. This process involves all of the activities listed under 'article creation costs' in the table above. This aspect of the economics of journal publishing is important as it is widely reported in the literature that these particular costs will not necessarily be impacted by any change in journal publishing business model, revenue source or distribution medium, i.e. it is implied that they are common to all models (see also: Area 6).

The table below presents data from a range of sources relating to these 'first copy costs'. The data collected by the two primary studies, Dryburgh (Source 2.2) and Waltham (Source 2.10), both show considerable variation between responses, but their ranges are broadly comparable ($310-$1,870 compared with $400-$1,600), as are the median values derived from the data ($700 compared with $915).

Table 2.2: Article creation ('first copy') costs per published paper (US$)

Data source	Date	Sample	Estimated article creation costs per published paper
Tenopir & King	1995	Analysis of three primary studies from 1970s and updated to 1995 values	$1,700
Dryburgh	2002	Primary data from 10 publishers (8 not-for-profit); 9/10 publish in science, range in output from 1 title to 140 (thus primarily small to medium-sized publishers covered)	***Range: $310-$1,870 Median: $700
SQW	2004	Review of primary literature/discussions with publishers	Range: $250-$2,000 *Good to high quality journal: $1,650 *Medium quality journal: $825
SQW	2004	Case study: large commercial publisher, large (in terms of extent), high quality journal	$2,200
SQW	2004	Case study: large commercial publisher, smaller (in terms of extent), medium quality journal	$350
Waltham	2005	Primary data collected in 2004 from seven publishers (all not-for-profit); cost data points collected for 12 journals; all STM journals; one journal in the sample is published using author-pays model and is online only	**Range: $400-$1,600 Median: $915 Mean: $1,000

* SQW defines a 'high quality journal' as having an 80-90 per cent rejection rate and a 'medium quality journal' as having a 40-50 per cent rejection rate

** Converted as per exchange rate used in the source report: $1.88=£1.00

*** Converted as per exchange rate quoted for September 2002 (US Federal Reserve): $1.56=£1

Article rejection rates

Anonymous case study findings quoted in the SQW report (Source 6.7) and tabulated above appear to confirm further how journal first copy costs can vary dramatically. The quality and volume of illustrations and graphics can vary by journal (and by discipline), as can the quality of the writing (editorial costs can be higher, for example, when submissions are from non-native English language speakers). Authors on the topic agree that variation in first copy costs is also attributable to the volume of manuscript submissions received and rejection rates applied for different journals (and the concomitant costs in editorial and peer review activities).

"Each article submitted must go through the peer review process and where large numbers of articles are rejected this, in effect, adds to the first-copy cost of those accepted" (SQW: Source 6.7)

The available evidence suggests that article acceptance and rejection rates vary considerably across journals. The data showing diversity in article rejection rates supports the evidence that per article costs can vary considerably between scholarly journals.

Table 2.3: Examples of manuscript rejection rates

Data source	Date	Sample	Rejection rates
Mellman	2004	Case study: Journal of Cell Biology	80-85%
Waltham	2005	Primary data collected in 2004, rejection rate data relates to eleven journals; all STM journals; one journal in the sample is published using author-pays model and is online only	Range: 42%-72% Median: 57%

Peer review

Dryburgh (Source 2.2) provides us with the only primary evidence relating to the operational costs of peer review to publishers. These costs embrace managing the receipt, review and processing of each manuscript submitted to a particular journal, including article administration and version control, but excludes copy-editing and re-writing activities. Estimates of staff salary costs in Dryburgh are augmented in all cases by 50 per cent to cover overheads, such as employers National Insurance, pension contribution, rent, rates and so on. This rule is applied somewhat arbitrarily to all cases and without explanation. No mention is made of whether any electronic peer review management systems were operated by responding publishers. Dryburgh's data (underpinned by seven valid responses) showed peer review costs per paper ranging from circa £30-£150, with a median value of £75. As noted above, primary analysis of peer review costs from Donovan (1998) was disregarded owing to the lack of transparency regarding methods and scope. Secondary analysis by Rowland (2002) has very little robust data on which to draw and is likewise not included in our analysis. Unfortunately Waltham (Source 2.10) does not separate out peer review costs from overall content creation costs in her analysis. As a result, data in this area are very sparse and cannot be satisfactorily cross-validated.

Total publishing costs

In addition to content creation costs, publishers incur both variable costs relating to production and distribution and also general overheads. Variable costs for journal publishing are very difficult to approximate as they vary significantly according to:

- Nature of the content (illustrations, colour, graphics can all impact on costs);
- Journal extent (number of articles, pages and issues);
- Distribution medium (print *vs.* online *vs.* print *and* online *vs.* mobile versions *vs.* podcasts and so on); and
- Circulation (size of the subscriber base in terms of both distribution and customers to be serviced).

It should also be noted that not all variable costs are linear (for example, if a publisher prints half as many journals, the costs will not halve). Thus, the impact of volume change can also vary by process.

SQW (Source 6.7) interpret Dryburgh's (Source 2.2) data to propose that 'first-copy costs'/content creation costs account for around 55 per cent of total costs (excluding overheads), other fixed costs account for around 5 per cent and variable costs generally account for around 40 per cent (again excluding overheads). However, this estimate does not correlate well with the comparable data derived from Waltham (Source 2.10). This latter survey provides detailed total cost breakdowns for 12 learned society journals. The survey findings demonstrate variable cost elements, representing a rather higher proportion of

overall costs (again excluding overheads) than SQW estimate. The variation in terms of range of responses indicates that a 'broad-brush' estimate of either variable or content creation costs probably oversimplifies the picture too much to be useful and that the balance of content creation versus variable production and distribution costs will vary widely from journal to journal according to diversity in terms of journal scale (extent and frequency), size of subscriber base and the nature of its content (e.g. number of illustrations).

Table 2.4: Content creation costs and variable costs of journal publishing (as % of total costs excluding overheads)

	SQW (Estimate based on Dryburgh data)	**Waltham (Mean values expressed as %)**	**Waltham (Ranges)**
Content creation costs	55%	46%	23%-69%
Other fixed costs	5%	-	-
Variable costs	40%	64%	49%-77%
Total (excluding overheads)	100%	100%	-

[Source: Waltham and SQW]

Tenopir and King's 2000 model provides a proportionate breakdown of total costs (including overheads) which correlates well with Waltham, with the exception that the proportions for manufacturing costs are lower in the older figures, although Tenopir and King's estimate easily falls within the overall ranges quoted in Waltham.

Table 2.5: Breakdown of total journal publishing costs (%)

	Tenopir & King 2000 (Average scholarly journal)	**Waltham 2004 (STM journals)**	**Waltham 2004 (Ranges)**
Content creation	37%	35%	16%-54%
Overheads/publishing support costs	30%	24%	11%-55%
Manufacturing, printing and paper	19%	28%	8%-40%
Distribution and fulfilment	14%	13%	3%-17%
Total	100%	100%	-

[Source: Waltham]

While the Waltham data are the most detailed available, providing a view of both variable costs and overheads per journal, per article and per page, they provide an incomplete picture. Circulation figures would have aided interpretation of the data, but are not available. The sample respondents are primarily UK-based, are active only in STM segments and are all learned society publishers. Clearly cost data from commercial publishers and university presses, and also journal publishers in social sciences, arts and humanities would be required to provide a more broad-based picture. Equally a larger volume of respondents would be desirable to identify meaningful trends amongst these sub-groups (for example, by type of publisher, size of publisher and by discipline) in a market universe in which variation is significant.

Libraries

For the purposes of this study, we have defined journal supply-side economics as the costs and effort incurred through the process of creating, producing and distributing scholarly journals to end-users (i.e. the research community). As such we embrace within our definition of supply-side economics the costs incurred by libraries in supplying content to end-users, i.e. the non-subscription related costs to libraries of the current journals system (as distinct from the

actual acquisition costs of subscriptions to journals, which is covered in Area 1 of this study). We have identified one major source in this area from Schonfeld *et al.* (Source 2.5). The report contains data on the non-subscription related costs of 11 US academic libraries (categorized as small, medium and large) and cross-compares relative costs of print holdings versus electronic periodicals, both for the year of data collection (2003) and modelled forward over a 25-year period.

The authors found that non-subscription costs[67] to libraries are not insignificant and that on a per-title basis costs for electronic formats (on average $21 per title) are consistently less than for traditional print titles (on average $89 per title). They conclude that:

"The transition to electronic periodical formats is bringing with it changes in library operations that will afford reductions in libraries' long-term commitments to non-subscription costs". (Schonfeld *et al.*: Source 2.5).

They recognize, however, that the transition to electronic formats is leading to larger collection sizes at libraries and that this will impact on per-title costs, especially for smaller libraries and furthermore that libraries face management challenges in the short-term. It is also recognised that the validity of their dollar-to-dollar print-to-electronic comparison is undermined by some system-wide shifts in costs. One important issue is the fact that the cost of day-to-day storage of periodicals, borne by librarians in the print format, is more often borne by publishers for electronic periodicals (in terms of providing server space and access), as shown on the table above describing the allocation of various costs.

Unfortunately, no comparable UK primary data on non-subscription costs specifically relating to journals could be identified, nor any data relating to the corporate library environment. There is no evidence against which to cross-validate these findings.

Differences between the economics of different types of publishers: non-profit distributing and commercial journal publishers

A cross-comparison of the broader economics of different types of journal publishers would ideally look at a range of publishers of different size and discipline focus in each of five main groups:

1. Commercial publisher
   and, non-profit distributing players:
2. University press
3. Learned society (in-house publisher)
4. Learned society (outsource publishing to commercial publisher)
5. Other: professional associations, intergovernmental organisations research foundations, charities (in-house publisher)

A review would address issues of costs, revenue and profits/surplus to identify any significant differences in the economics of journal publishing between these types of operator. For example, an analysis would look at the breakdown in revenue generated by different customer segments (e.g. member subscriptions *vs.* institutional subscriptions *vs.* individual subscriptions) and the distribution of revenues by type of content sale (e.g. subscriptions *vs.* author payments *vs.* individual article sales) and so on. However, once again, the existing evidence

[67] Non-subscription costs were assigned under 14 main categories: collections development, licensing and negotiations, subscriptions processing, routine renewals and termination, receipt and check-in, routing of issues and tables of contents, cataloguing, linking services, physical processing, stacks maintenance, circulation, reference and research, user instruction, preservation, other.

base is insufficient to permit this type of analysis. The most useful primary research available relates almost exclusively to the learned society segment of the market; as has been noted already, in-depth economic analyses of journal-level publishing in the commercial segment is a key gap.

Table 2.6: Matrix of existing evidence base on economics of journal publishers by type

	Costs	**Revenues**	**Profit/surplus**
Commercial publisher	No evidence available (see previous section)	EPS: Informed top-level estimates for largest c10 STM players available from Market Monitor (2006). No data available at the individual journal level or for publishers in non-STM disciplines. Case studies from individual publishers.	EPS: Secondary data from published sources for top c10 STM players available from Market Monitor (2006) No data available at the individual journal level. or for publishers in non-STM disciplines. or for publishers in non-STM disciplines. Case studies from individual publishers.
*University press**	No evidence available	No evidence available	No evidence available
Learned society (in-housel)	Waltham: Primary survey of thirteen journals; detailed journal cost breakdown (2005)	Waltham: Primary survey of thirteen journals; analysis of revenue sources, levels and trends by journal (2005)	Waltham: Primary survey of thirteen journals; range of surplus values for respondents by journal (2005)
Learned society (contracted-out)	No evidence available	No evidence available	Baldwin: Survey regarding surplus revenues of society publishers and what they use it for; responses from 68 society publishers; 44% of which do their own publishing, 43% contract out and 13% do both. (2004)
Other non-profit distributing publishers	No evidence available	No evidence available	No evidence available

*[*Whilst two university presses are included in Dryburgh's review, the data are anonymised so cannot be analysed separately]*

Costs of launching new journals

Our literature review revealed no primary source material relating to launching new scholarly journals. A useful, if slightly dated, summary is given in Page *et al.* (Source 2.4). The authors also refer to Bailey,[68] who discusses how publishers decide whether or not to publish a new journal. Page *et al.* (Source 2.4) estimate a journal launch as costing 'tens of thousands of dollars or pounds over several years,' with a high risk of failure. They suggest that it may take up to five years for a new journal to break even and even longer to pay back the initial investment – typically a publisher will produce a five-year plan as part of a new journal launch. Initial set up costs (with the process usually beginning six months before first issue publication) are described as publicity (attracting both authors and subscribers), producing sample copies and setting up an editorial

[68] Bailey, J (1989). New Journal Decision Making, College & Research Libraries, May, 1989.

office. Publishers can opt for a low-cost strategy (minimal publicity, low fee for editor, lower quality production). A high-cost strategy is estimated to result in a deficit of $150,000 by end of year three of publication (promotional costs for the first 18 months in a new journal's life-cycle are estimated at $30,000). In some cases if a new title has a sponsor (such as a scientific society) both the risks and costs may be less burdensome (if, for example, a society's membership base is willing to pay above cost to receive the new title). Many libraries take three years or more to allocate funds to acquire a new journal so subscriber growth is usually gradual. The third year is often a critical milestone in terms of both sufficient subscriber growth and high quality article submissions to enable publishers to assess the title's likely success or failure. However, no primary evidence is produced to support these arguments or to take account of potential diversity of practice and investment levels in the market.

Gaps in the data

We have some helpful base data on costs across the various participants in journal publishing. There are gaps and weaknesses, but in the area of supply-side economics many of these can be overcome through further research. Firstly, the data relating to the costs incurred by researchers are dated, not UK-specific and pre-date the widespread introduction of electronic journals. Further primary research work could be undertaken here to quantify their contribution. It should also be possible to collect cost (and potentially revenue) data anonymously from commercial and non-profit distributing publishers via a primary survey (similar to those already undertaken in the learned society arena) both to improve the overall evidence base regarding the costs of the existing journal system (including launching new journals) and to identify and cross-compare the economics of different types of players and the relative costs of different elements in the publishing process. Gaining the trust of the players involved and ensuring participation of a sufficiently wide sample of players to provide a broad and representative picture across types of publisher and journal would be the key success factors. The evidence of diversity in journal publishing economics that already exists in Waltham (Source 2.10) and Dryburgh (Source 2.2) suggests that this issue will be more productively investigated at the journal level, and that data giving ranges of costs by discipline, frequency, extent and circulation will most accurately reflect the true complexity of supply-side costs.

Sources evaluated in detail in Appendix 1

Source 1: Baldwin, C (2004). What do Societies do with their Surpluses? ALPSP/Blackwell Survey

Source 2: Dryburgh, A (2002). The costs of learned journal and book publishing: a benchmarking study for ALPSP.

Source 3: Morris, Sally (2005). The True Costs of Scholarly Publishing. Learned Publishing, 18(2), 115-126.

Source 4: Page, G, Campbell, R, Meadows, J (1997). Journal Publishing. Cambridge University Press, pp 407 ISBN 0 521 44137 4. Chapter 10: Managing a List of Journals

Source 5: Schonfeld, King, Okersen and Fenton (2004) on behalf of the Council on Library and Information Resources. The Non-Subscription Side of Periodicals: Changes in Library Operations and Costs between Print and Electronic Formats.

Source 6: Tenopir & King (December 1998). Designing Electronic Journals with Thirty Years of Lessons from Print, The Journal of Electronic Publishing, Volume 4, Issue 2

Source 7: Tenopir & King (31 March 1998) Economic Cost Models of Scientific Scholarly Journals, University of Tennessee School of Information Sciences Center for Information Studies Paper presented to the ICSU Press Workshop, Keble College, Oxford, UK

Source 8: Tenopir, C and King, D (2000). Towards Electronic Journals: Realities for Scientists, Librarians, and Publishers. Washington, D.C.: Special Libraries Association.

Source 9: Tenopir & King, (2004). An evidence-based assessment of the 'author pays' model, *Nature* Web Focus, June 2004, http://www.nature.com/nature/focus/accessdebate/26.html

Source 10: Waltham, M, on behalf of JISC (June 2005). Learned Society Open Access Business Models.

Area 3: Usage

Executive summary

Key questions

This review of journal usage focuses on evidence relating to the use of journals by both readers and contributors. Key questions identified were:

1. How much use is made of journal articles?
2. What is the split between leading journals and the rest?
3. What is the extent of unmet demand for journals?
4. What is the potential for this demand?
5. What are the barriers to this demand being met?

The evidence

The evidence base on scholarly journal usage shows a good deal of consensus on key issues. However it is incomplete from a couple of perspectives. Firstly, the primary data focus on authors and users in higher education – corporate and other end-users are under-represented. Furthermore, much of the available evidence is based on user surveys rather than actual usage data (i.e. it records what users *say* they do, but we can not cross-validate this with actual *behaviour*). Valuable work has been undertaken in analysing transaction logs for electronic journal collections, but little of the data is currently in the public domain.

Readers value journals and particularly welcome the flexible access afforded by electronic content (although printing is still common practice for reading purposes). Journal reputation (and the peer review process underlying this) continues to be valued by readers and authors alike. Maintaining the peer review system is seen as important by all constituents of the user community, regardless of the publishing model used.

Surveys record that researchers can experience access problems (e.g. the library not stocking a particular journal), but these are not reported to be a major obstacle to research productivity overall. There is evidence that electronic resources are improving access. Interdisciplinary researchers are highlighted as facing the most difficulty in both accessing journals and in finding publishing outlets.

The gaps

Publicly available hard data on journal usage at the article level (in either hard or electronic copy) constitute a key gap. For example, it is possible that data logs showing unsuccessful requests for journal access would provide useful quantitative evidence of unmet demand (by readers).

Increasing electronic access to journals facilitates the collection of actual usage data and this resource will be critical to cross-validate and enrich data collected through survey methods.

There is no evidence relating to the precise nature of the use made of the information in journals that are accessed. We have no detailed view as to how access to journals 'adds value' to researchers' endeavours and productivity. This, in the context of emerging alternative modes of scholarly communication, is a significant gap to be filled by appropriate research.

Introduction

This review of journal usage focuses on evidence relating to the use of journals both by readers and contributors, including any evidence for unmet demand. Key questions identified were:

1. How much use is made of journal articles?
2. What is the split between leading journals and the rest?
3. What is the extent of unmet demand for journals?
4. What is the potential for this demand?
5. What are the barriers to this demand being met?

Several of these questions are open to various interpretations, as discussed below. Together with the type of data available, this ambiguity means that clear answers are not always possible. However, the evidence identified does give insight into readers' and contributors' perceptions of the scholarly publishing process, with little contradiction.

This Area should also be read in conjunction with Area 5 'Disciplinary differences'. Since the sources for Areas 3 and 5 overlap to a very considerable extent, they are presented in a single section in Appendix 1 of this report.

Scope and definitions

Having initially reviewed the key sources, it was clear that the issues enumerated above required further clarification, and we have outlined our interpretation of each below. Not all points were covered explicitly in the sources, but the data may indicate a trend. In general, surveys provide the richest data, but the focus, population samples, and methods of individual studies, together with the form and content of survey questions, obviously differ, so direct comparison of results should be viewed with some caution.

Question 1: This question is approached from both a reader and contributor perspective, but it is not always possible to drill down to article level, nor to specify precisely how a reader 'uses' an electronic source (e.g. downloads full article, browses the contents pages etc.) as these data were not collected by the original researchers. The level of analysis is reported as it is given in the sources. Any indication of use is included in the source reviews, but not all of these can be used for triangulation as not all sources address the same questions.

Question 2: In the absence of any definition of 'leading', the reviewers have commented on high reputation/impact factor journals, as this issue emerges as important to journal users. For cross comparison questions 1 and 2 have been merged.

Question 3: This question has also been addressed from the perspective of readers and contributors. The literature indicates that from a reader's point of view 'unmet demand' usually means issues related to 'access' (e.g. from a library), rather than there being a demand for a journal product which does not currently exist anywhere. As a result, this latter aspect is not directly covered, but certain assumptions can be made for both readers' and contributors' demands.

Question 4: The lack of firm evidence of any unmet demand means that this question cannot be answered. For cross comparison questions 3 and 4 have been merged.

Question 5: Evidence of barriers is limited (as there is no real evidence of unmet demand). Any barriers identified will also be discussed under Question 3.

Most of the data relate to use by academics in HEIs. The journal usage of researchers in other environments (such as private enterprise or government organisations) is poorly covered. Most of the surveys were either UK based or international in scope. Any exceptions are noted.

Key sources

None of the key sources provides suitable data for answering all of the questions concerned with usage. The sources provide one of two distinct types of data to report on aspects of scholarly communication.

Transaction logs

Part of the ongoing CIBER work is analysis of transaction logs, by a methodology called 'deep log analysis' (Sources 3/5.7 and 3/5.8). This aims to reveal user behaviour as well as more usual usage statistics.

This research is part of the CIBER team's Virtual Scholar Research Programme which has produced a series of published papers communicating findings from innovative analysis of usage log data for electronic journals. The team has also published articles and reports of its extensive author surveys (described below), for which bespoke analyses of the anonymised data set are also available. In terms of papers reporting transaction log analyses, we have drawn mainly on two sources (see also Appendix):

- ✓ 3/5.7 *Nicholas, D. et al (2005), The big deal – ten years on, Learned publishing 18(4) October 2005* – an exploratory deep log analysis of the OhioLINK journal set; and,
- ✓ 3/5.8 *Nicholas, D., Huntington, P. and Watkinson, A. (2005). Scholarly journal usage: the results of deep log analysis. Journal of documentation 61(2) pp.248-280* – a study of Blackwell Synergy users.

As an example, the latter study provides objective usage metrics for nearly one million users (and combines this with user demographic data for a sub-set of 500,000 of these). The insights provided regarding information-seeking behaviour provide a uniquely detailed and objective view and include useful additional analyses by broad subject grouping (science vs. medical vs. academic vs. professional) and by occupation (professor/teacher vs. researcher vs. post-graduate vs. under-graduate vs. professional). These two sources are the most relevant to the particular questions addressed in this study (and are judged to be broadly indicative of the wide range of CIBER literature in the public domain) and, hence, these two papers have been selected as the focus of attention in this report. The results reported represent a selection of findings from the underlying transaction logs and concentrate on search behaviour in higher education environments. However, the data collected could be used to answer some of the questions relevant to Areas 3 and 5 of this project.

Evidence Base (Source 3/5.1) analysed usage statistics in a selection of UK HEIs in order to inform licence negotiations. Unfortunately, the full data set is

confidential, but a summary report indicates that useful data have been collected. For example, it was found that a small percentage of journal titles accounted for a high amount of usage and that usage was highest in the high/very high price bands. Free-of-charge or low-priced journals accounted for nil or low usages. These data support the findings from the author surveys reviewed below. Other data collected by Evidence Base included COUNTER reports which give, among other information, details of full-text downloads and unsuccessful attempts at downloading. These data would provide useful data relating to both usage patterns and unmet demand. However, these data are confidential and are not published.

Surveys

These surveys usually take the form of web-based questionnaires sent to 'published' authors. Most prominent in this field is the work by CIBER at UCL; their near annual surveys of authors are reported in multiple papers, each with a slightly different emphasis. However, the available data has two main limitations. Firstly, the primary focus of the CIBER research team is open access and this can limit the applicability of its findings. Also, although a considerable amount of data have been collected, and some are relevant to this area of the project, the papers reviewed present only a selected sub-set of the data, as most of the original research is undertaken confidentially on behalf of publishers and cannot be made public. Key Perspectives Ltd have also undertaken valuable author surveys for ALPSP and these are described below. Sources reviewed for this Area are listed at the end of this section and full evaluation sheets are provided in Appendix 1. Certain surveys were selected for more detailed discussion of results (see Analysis). The nature of these primary surveys is such that they often provide an array of subjective views from respondents, and while some inferences may be drawn from the data it would be unwise to use them to arrive at firm conclusions concerning the overall picture relating to access issues.

Source 1 is discussed above in 'transaction logs'.

Source 2: Education for Change, SIRU, and Research Partnership. Researchers use of libraries and other information sources, 2002

This is a survey of UK researchers' use of libraries and therefore much is not relevant to this area of the project. However, it does give an indication of the importance of printed journals at the time of the survey and the anticipated role of electronic journals in the future. Questionnaires (postal/online) were sent to 3,390 researchers in HE – sampled from the RAE census. The survey achieved a 45 per cent response which was representative of the sample. The report looks at researchers' use of all materials, so puts journal use in wider context.

Source 3: eJust (e-Journal user study). Stanford University, 2002.

Multi-part study based on use of e-journals. It used qualitative and quantitative methods, including log data. The study focused on users of Highwire Press Journals in the life sciences who were mostly based in the USA (though with some representation of other countries). The web site provides reports relating to each stage of the project, plus full results linked to individual questions.

Source 4: Key Perspectives Ltd for JISC/OSI (2004). JISC/OSI Journal authors survey.

Survey to compare the experiences of authors who had published in open access journals with those who had not. The survey had a UK focus. Data were collected using two web-based questionnaires – one for OA authors, one for non-OA authors. These were sent to 3,059 OA authors and 5,000 non-OA authors;

response rates were 5 per cent (154) and 3 per cent (157) respectively. The general focus on OA *vs.* non-OA author views is of limited relevance to the present study, but this survey does give some indication of future preferences.

Source 5: Key Perspectives Ltd, (1999), What authors want: the ALPSP research study on the motivations and concerns of contributors to learned journals.

This is a survey of authors of papers published in scholarly journals. Questionnaires were sent to 10,970 authors who had contributed to a selection of journals. All disciplines were included and the survey was international in scope. A total of 3,218 responses were received. Although old now this study covers key questions about authors' involvement in the publishing process. Its scope is mainly restricted to print journals; its questions about the likely impact of e-journals are of particular interest when compared with later studies.

Source 6 was included for review, but not covered here as it is an interesting paper but outside the main argument.

Sources 7 and 8 are discussed above in 'transaction logs'.

Source 9: Rightscom Ltd. (2005). JISC Disciplinary differences report.

This is a web survey of 780 academics in UK Higher Education Institutions to inform JISC on information needs of researchers and the barriers and opportunities facing researchers both in access to research resources and in publishing their work. The survey covers a range of information resources, not just journals. The main focus of the research is disciplinary differences in use of resources (see also Area 5), but it does report on some issues of general interest, particularly regarding access to materials. The survey achieved 780 responses covering all RAE units of assessment (except Celtic Studies), although some areas are somewhat under-represented.

Source 10: Rowlands, I. and Nicholas, D. (2005). New Journal Publishing Models.

Part of an ongoing study into changing nature of scholarly communication, this paper is the latest in series focusing on authors' views, focusing particularly on open access. This was a web survey of 76,790 researchers who had published in an ISI indexed journal during 2004 and it achieved a 7.2 per cent response rate.

Source 11: Rowlands, I. and Nicholas, D. and Huntingdon, P., (2004). Scholarly communication in the digital environment: what do authors want?

The focus of this survey was usage of and awareness of open access, which is reflected throughout the report. The web-based questionnaire used 'critical incident' technique – i.e. answers are based on last published article. Questionnaires were distributed to 107,500 authors who had published in a peer-reviewed journal in the last 18 months. The study covered 97 countries. The response rate was 4 per cent. The sample was derived from the ISI mailing list, and respondents, therefore, tended to be senior/experienced researchers.

Source 12: Rowlands, I. and Olivieri, R., (2006). Journals and scientific productivity.

This study examines attitudes to the journal system and behaviour of researchers in immunology and microbiology – both as authors and readers. Includes analysis of 2004 and 2005 CIBER surveys, plus a survey of 883 researchers in these disciplines. This report considers journals as just one part of whole 'research environment' to look very broadly at barriers to researcher productivity.

Source 13: Key Perspectives Ltd, (2002). Authors and electronic publishing: the ALPSP research study on authors' and readers' views of electronic research communications.

This is a web-based survey of 14,643 authors. Authors were solicited by approaching publishers for details of contributors to a selection of their journals. Twenty-three publishers participated, including society, university, and commercial players. The response rate was 8.5 per cent (1,246 responses). Full results are provided in the report, including breakdown by discipline, making this a useful source.

Source 14: Tenopir, C et al. (2003). Patterns of journal use by scientists through three evolutionary phases

This secondary source provides a review of US-based authors' surveys of e-journal usage since 1990. Patterns of usage are divided into three phases: early (1990-1993) based on a surveys of 862 scientists in two US universities and eight other research organisations; evolving (late 1990s-2002) based on a survey of 235 scientists in two US universities in 2002; and an advanced stage. The advanced phase is based mostly on future scenarios (for mainstream use) based on a system developed by the American Astronomical Society, and the survey is limited to their members in 2002 (508 responses). As the samples and questions differ between survey, firm conclusions cannot be drawn. Particular caution is needed for interpretation of the 'advanced stage' as this is limited to a small section of the scientific community and their use of a specialised system.

Analysis/data validation

Data validation is difficult because the individual studies do not lend themselves to detailed comparison. The character and execution of surveys differ materially, as noted earlier. The majority of surveys concentrate on subjective views of the situation by academics, for example by asking for their degree of agreement with certain statements. Therefore, it is only possible to discuss broad agreement between survey results and much of the character and richness of original surveys is lost in the process. It should be noted, however, that this top-level cross-comparison reveals a good degree of consensus between sources – there do not appear to be any contradictory data in relation to any of the major issues.

How much use is made of journal articles?

Use by readers

Journal articles are viewed as an essential research resource by the reader community (Sources 3/5.2, 3/5.3, 3/5.9). Although there are no hard data on the number of articles actually read, the surveys suggest that journal usage is generally increasing. Tenopir (Source 3/5.14) reviewed reader surveys carried out over a period of 13 years and found the number of articles read was increasing, but the reasons for valuing journals as a resource remained the same. Swan (Source 3/5.13) found that readers valued the increased accessibility of electronic journals, and articles were accessed from home during evenings, weekends and holidays as well as during working hours.

It should be noted again that further transaction log data on usage are collected by individual institutions and publishers (and would offer direct evidence of behaviour to cross-compare with survey responses), but that these data are not publicly available.

All sources agree that the reputation of journals is of key importance and that some form of peer review should be maintained. Although there is general support for electronic journals (Sources 3/5.2, 3/5.3, 3/5.11, 3/5.13, and 3/5.14 in particular), users tend to print articles to read (Sources 3/5.3, 3/5.9), and future models tend to include print sources (Sources 3/5.3, 3/5.13).

Use by authors

Journals are an important dissemination source for academics (Source 3/5.9). There is a very strong preference for peer-reviewed journals, and all sources found evidence that the reputation of a journal (including impact factor) was the most important feature when authors sought an outlet for articles. This was further supported in authors' opinions of future models. The Key Perspectives study (Source 3/5.5) noted that authors would like a future model which included electronic journals, but favoured maintenance of a peer review system. A later study (Source 3/5.4) supports this – authors would welcome more open access journals, but stressed that peer review must be maintained.

Unmet demand on the part of readers

The issue of access to resources is covered in surveys in various ways. Access problems were not found to be a major obstacle to research productivity (Source 3/5.11). There is also evidence that electronic resources are improving access (Sources 3/5.11, 3/5.12). The Rightscom survey (Source 3/5.9) found that just under half of respondents had experienced problems accessing research resources. This covered all resources, but a particular problem is the library not stocking a particular journal. Source 3/5.13 presented some evidence that researchers will try to obtain articles that are not immediately available; indeed, only 25 per cent never bothered to obtain articles from other sources. Two surveys found that access to journals was more difficult for interdisciplinary workers (sources 3/5.9, 3/5.12).

The eJust survey (Source 3/5.3) found that readers are unwilling to pay for one-off articles, but that there is some indication that online access motivates users to take out personal subscriptions or join societies. This is not supported by Tenopir (Source 3/5.14), who found that the number of personal subscriptions is decreasing, while use of library collections is increasing. The largest increase found in this review of surveys since 1990 was in the use of 'separate copies', including preprints, inter-library loans and author websites.

Unmet demand on the part of authors

This issue is not directly addressed by the studies reviewed. However, barriers to publishing cited by authors included publication delays (Sources 3/5.5, 3/5.9) and the peer review process (Source 3/5.5). However, the problems with peer review were related to publication delay.

There is limited evidence on the difficulty or otherwise of finding a suitable outlet for work. Although the Rightscom survey (Source 3/5.9) reported problems in placing articles in high-quality journals, another survey noted that researchers think that too much is being published (Source 3/5.10). Surveys support the view that authors prefer to place articles in high-quality journals, not least because this is important for career development. Therefore, demand for these journals is high, but there is no evidence that there are insufficient publishing outlets. However, the exception is interdisciplinary researchers, who are reported to have more difficulty in getting work published owing to lack of appropriate outlets (Source 3/5.12).

Gaps in the data

How much use is made of journal articles?

The value of journal articles is well documented in the surveys reviewed. However, usage statistics to an article level are not readily available; although these statistics are collected by various publishers and individual institutions, they are not publicly available.

The original research brief required evidence relating to journal usage to be segmented between 'leading journals' and 'the rest'. To discuss this issue a clear definition of 'leading' is required, and the same problems with lack of available hard data arises as for general use above.

Another issue relates to the general gap in our knowledge regarding the precise nature of the use made of the information in journals that are accessed. How does access to journals 'add value' to researchers' endeavours? This, in the context of alternative modes of scholarly communication, is a significant gap to be filled by appropriate research.

What is the extent of unmet demand for journals?

It is possible that data logs can show unsuccessful requests for journal access, which would provide useful quantitative evidence of unmet demand (by readers) within institutions. A component of these unsuccessful requests may, of course, include instances in which, for example, network congestion renders initial attempts at access unsuccessful, followed by successful episodes. Detailed analysis of the data could reveal the extent of delays and difficulties that users encounter in this way. However, none of this type of data is publicly available at this time. Author/reader surveys do cover the issue of ease of access to journals to some extent, but the question is raised in different ways so it is difficult to get a clear picture. An analysis of evidence relating to trends in inter-library loans would also be useful here to examine the extent to which readers appear to continue to rely on this access method to obtain materials not available at their own libraries. Most data relating to access concentrate on accessing articles readers know exist. The more nebulous question of an unmet demand for a particular type of journal is not covered in the literature. The issue of unmet demand on the part of authors is well covered by surveys.

Sources evaluated in detail in Appendix 1

Source 1	Conyers, A. and Dalton, P., Evidence Base, (2005). NESLi2 analysis of usage statistics: summary report.
Source 2	Education for Change, SIRU, and Research Partnership, (2002). Researchers use of libraries and other information sources.
Source 3	eJust (e-Journal user study), (2002). Stanford University.
Source 4	Key Perspectives Ltd. for JISC/OSI, (2004). JISC/OSI Journal authors survey, 2004.

Source 5	Key Perspectives (1999). What authors want: the ALPSP research study on the motivations and concerns of contributors to learned journals.
Source 6	Loke, Y.K. and Derry, S. (2003). Does anybody read 'evidence based' articles?
Source 7 2005	Nicholas, D. et al., (2005). The big deal – 10 years on.
Source 8	Nicholas, D et al., (2005). Scholarly journal usage: the results of deep log analysis.
Source 9	Rightscom Ltd., (2005). JISC Disciplinary differences report, 2005.
Source 10	Rowlands, I. and Nicholas, D., (2005). New Journal Publishing Models.
Source 11	Rowlands, I., Nicholas, D., and Huntingdon, P., (2004). Scholarly communication in the digital environment: what do authors want?
Source 12	Rowlands, I and Olivieri, R., (2006). Journals and scientific productivity.
Source 13	Key Perspectives Ltd., (2002). Authors and electronic publishing: the ALPSP research study on authors and readers views of electronic research communications.
Source 14	Tenopir, C et al. (2003). Patterns of journal use by scientists through three evolutionary phases.

Area 4: Citations, Impact Factors and their role

Executive summary

Key questions

The three questions considered in this area of the report are:

1. Are traditional (i.e., subscription-based) journals more likely to be cited than OA journals, and as a result, do impact factors (IF) vary between traditional and OA journals?
2. What impact do citations and IFs have on research funding?
3. How are IFs affected by the increasing trend to collaborative research and multi-location research?

The evidence

This is an area in which much research has been carried out, but most of it has been on specific subject areas or titles, making it difficult to generalise. Furthermore, any study of variance in article impact in OA environments versus subscription environments faces a key methodological challenge in that a given article cannot be OA and non-OA at the same time and, therefore, an exact like-for-like comparison of research impact over the same time period is not possible.

Much of the research that has been undertaken in this area has been on self-archived articles as opposed to articles that have appeared in OA journals; it is important to distinguish carefully between the two when interpreting the findings.

There is some consistency in results that show more citations for articles self-archived in repositories as distinct from the same or similar articles available in a subscription journal (although there have also been a few contradictory results). Overall, deposit of articles in open access repositories seems to be associated with both a larger number of citations, and earlier citations for the items deposited.

The reasons for this, however, have not been clearly established - there are many factors that influence citation rates, including the reputation of the author, the subject-matter of the article, the self-citation rate, and, of course, how important or influential the repository is in its own right. The little existing evidence suggests that a possible reason for increased citation counts is not that the materials were free, or that they appeared more rapidly, but that authors put their *best* work into OA format. This research was limited to one discipline, however, and more extensive evidence is required to validate this finding.

There is less consistent evidence relating to IF advantage for OA journals over toll-access journals, although OA articles in a non-randomised study in a hybrid journal have been recorded as achieving higher citation counts (over the same time) than subscription-access articles in the same journal.

With reference to any impact of citations and IFs on research funding, anecdotal evidence suggests that some funding agencies use citation counts as part of the assessment procedure when allocating research funds, but the number of publications confirming this approach is tiny. There are more data available on the reverse scenario, i.e., the impact of funding or source of funds on citation counts. We conclude that the two are linked, but that only rarely are citation counts a key factor in funding decision-making.

In relation to any link between citations and collaborative research, there is some scattered, but consistent evidence in various bibliometric studies that multi-institutional or multi-nationally authored papers are more frequently cited than papers that come from a single institute. However, such increases in citation counts that do occur with regard to co-operation are likely to be a purely numerical artefact of the greater number of authors and can be explained by self-citation.

The gaps

Although quite a lot of evidence has been collected regarding the quantitative effect of OA on citation counts (whether in the form of OA journals or as self-archived articles), much of it is scattered, uses inconsistent methods and covers different subject areas.

Consistent longitudinal data over a period of years to measure IF trends in a representative range of journals would fill this gap, e.g., studying a range of journals that were toll-access and went OA (or *vice versa*).

In the short-term, more data in different disciplines measuring the impact on citation counts of articles in hybrid journals or articles that are available in both forms *versus* articles that are only available in one of the forms will improve the evidence base.

Qualitative factors should not be ignored, however, in either conducting or interpreting research on this issue. The existing sources rarely take into account the full range of factors that can affect citation counts. These are challenging data to collect, but not unprecedented.

Little or no research has been carried out on research funders' approaches to citations as a metric for decisions. A comprehensive review of policy statements coupled with a primary survey of funding agencies would be required to collect base data on this issue.

The current evidence base does not support the premise that collaborative research leads to a meaningful increase in citation counts (and, by implication, greater research 'success') once self-citation by a larger number of authors has been taken into account. It is clear that any new research seeking to validate this finding through impact measurement must involve corrections for self-citations.

Introduction

This review of citations and Impact Factors (IFs) focuses on the evidence that has been produced linking these metrics to whether source material is available in Open Access (OA) publications (i.e. OA journals and OA repositories) or in toll-access journals. This is an area in which much research has been carried out, but most of it has been on specific subject areas or titles, making it difficult to generalise.

The three questions considered are:

- Are traditional (i.e. subscription-based) journals more likely to be cited than OA journals, and as a result, do IFs vary between traditional and OA journals?
- What impact do citations and IFs have on research funding?
- How are IFs affected by the increasing trend to collaborative research and multi-location research?

Scope and definitions

For the purposes of this report, a "citation" is defined as "a reference to a document or part of a document from which a passage is quoted, or to a source regarded as an authority for a statement or proposition".69 Area 4 of this study seeks to identify evidence on citation patterns and, hence, IFs relating to subscription journals on the one hand and OA journals and repositories on the other. Definitions of both of these types of OA publishing vehicle are provided in Area 6 of this Report.

IFs have been used for many years in academic and scholarly publishing as a means of academic evaluation of researchers, as a journal collection management tool for librarians, and as a market research tool for publishers. The IF is defined by Thomson Scientific as a measure of the frequency with which the "average article" in a journal has been cited in a particular year or period. The annual IF is a ratio between citations and recent citable items published (it is calculated by dividing the number of current year citations to source items published in the journal during the previous two years).

It must be stressed that "Open Access" can take two different forms. The first form is self-archiving, whereby an individual places research output in a repository, which anyone with suitable hardware and software can access at no charge; the second form is by means of publication in an "Open Access Journal", i.e., an electronic journal for which there is no subscription charge. Some of the research that has been undertaken in this area has been on self-archived articles, and some on articles that have appeared in OA journals and it is important to distinguish between the two carefully when interpreting the findings

This report is not the place to discuss whether citation counts are a reliable method of assessing research quality. Despite a number of weaknesses associated with the technique, general opinion is of the view that high citation counts are associated with high quality and/or high impact of research output. The claim that negative citations will affect citation counts significantly, e.g., the debate about side-effects of MMR vaccines or the idea of cold fusion, and as a result invalidates citation analyses is oft-repeated, but in practice such perturbations are too small to disturb the general validity of the technique. A significant and widely-reported problem is that ISI's Web of Science is biased towards English-language publications, and is also weaker in its coverage of the arts and humanities than of science or even social science. For these reasons, results of citation analyses, especially analyses based purely on Web of Science, should be treated with caution.

[69] Taken from S. Keenan and C. Johnston, "Concise Dictionary of Library and Information Science", Bowker Saur, 2000. This definition is not perfect, as it takes no account of situations where someone cites a reference in order to criticise it, but has the advantage of being concise.

Key sources

The major sources of evidence on these issues are identified and summarised below. These key sources have also been evaluated to assess their applicability to the study. Individual source evaluation sheets for each are included in Appendix 1 of this report.

Are traditional journals more likely to be cited than OA journals?

There has been little primary research examining this area. The 2004 Thomson Scientific (formerly ISI) report, *The impact of open access journals*[70] (Source 1.13), is important, as it is both the most comprehensive of the sources identified and because it comes from the major citation database organisation. The authors find that citations from OA journals are not significantly different from those of non-OA journals; overall its findings provide little comfort to those who claim OA journals are more likely to be cited than subscription-based journals.

Bauer and Bakkalbasi[71] (Source 4.2) compare the citation counting ability of Web of Science, Google Scholar and Scopus in the context of changes to the scholarly publishing arena, and argue the case for *Google Scholar* for citation counting. However, Jacso (unpublished results) has prepared a devastating critique of the latter's accuracy, and it should not be relied upon. Other research (M. Norris and C. Oppenheim, paper in preparation for *Scientometrics*) has shown that Elsevier's *Scopus* is potentially a slightly better source of citation data than Thomson Scientific's database (*Web of Science)*. The current overall opinion is that *Web of Science* is the best source of citation data at the moment, but it is worth noting that it may not remain so in the future.

It is worth noting that there is no standard textbook or reference work on these topics, though Moed's book[72] provides an overview of the broader area. Shadbolt *et al.* provide a vision for the future of evaluation of research impact in an OA environment[73] and speculate about the types of metrics that might be used in the future.

Do IFs vary between traditional and OA journals?

Thomson Scientific's *Journal Citation Reports*[74] (Source 1.13) provide detailed IF data for thousands of journals, virtually all of which are peer-reviewed, but only a small minority of which are OA. The Thomson study (Source 1.13) described above showed little difference in IF between the two types of journals. This was a particularly wide-scale study using the best citation data currently available, and its results need to be taken seriously. The Oxford Journals study[75] considered in more detail below measures citation counts, rather than IF, so is not relevant here. Thus, the evidence for higher IFs for OA journals remains tenuous at the moment.

The link between Open Access and citation counts

---

[70] Thomson ISI, *The impact of open access journals – a citation study from Thomson ISI*, 2004, http://scientific.thomson.com/ts/media/presentrep/acropdf/impact-oa-journals.pdf
[71] K. Bauer and N. Bakkalbasi, An examination of citation counts in a new scholarly communication environment, *D-Lib Magazine*, 2005, **11** (9), http://www.dlib.org/dlib/september05/bauer/09bauer.html
[72] H.F. Moed, *Citation analysis in research evaluation*, Springer (2005)
[73] N. Shadbolt, T. Brody, L. Carr and S. Harnad, The open research web, in N. Jacobs (Editor), *Open Access: key strategic, technical and economic aspects,* Chandos Publishing (2006), 195-208.
[74] http://scientific.thomson.com/products/jcr/
[75] Oxford Journals, *Assessing the impact of Open Access*, Oxford University Press, 2006.

Self archived articles

It is this area that has been most studied, with numerous key publications. Most of these are focussed on the citation advantage of self-archived articles rather than of OA journals. Craig, in an as yet unpublished review[76] provides an excellent overview of the evidence collected to date. Lawrence[77] (Source 4.13) is significant because it was the first major paper that identified a citation advantage for OAself-archived articles, and it has been widely cited ever since. However, it was based on a too small-scale a study to support general conclusions. Harnad *et al.*[78] (Source 4.9) provides a useful summary of the state of play of OA advantage studies, while Hajjem *et al.* [79] (Source 4.8) is fairly typical of the many articles produced by Harnad claiming that self-archiving leads to higher citation counts. Antelman[80] (Source 4.1) is arguably the most carefully constructed study of the question. Articles in four disciplines were evaluated, and in each case it was found that open access articles had greater citation counts than non-open access articles.

Open access journals

Eysenbach's important article[81] (Source 4.7) presents the findings of a rigorous longitudinal study of citation counts for the journal *Proceedings of the National Academy of Sciences* (PNAS). PNAS is a "hybrid" journal, offering both OA articles and subscription-based articles in the same journal. The study compared citation counts at three different points in time for a sample of 1,492 articles (14% OA, 86% subscription access) adjusting results for particular author and article characteristics. The findings showed that OA papers in this title were more highly cited than non-OA articles in the same journal. This part of his paper was uncontroversial, but another part produced a heated debate about whether self-archiving or publication in an OA journal offers the best route to OA[82].

The Oxford Journals study[83] (Source 4.14) looks at the effect of the introduction of OA on downloads, citations and attitudes to a small range of Oxford Journals' life sciences journals. The method for the study is sound, being based on large samples of responses and deep log analyses. However, although some results are in accord with previous work (on downloads), the fact that citations for OA articles are lower than for subscription-based articles is surprising and counter-intuitive.

Most of these studies demonstrate a link between OA availability through either a journal or a repository and increased citation counts (and also speed of citation),

[76] I. Craig, The effect of self-archiving and Open Access on citations, Blackwell Scientific Internal Report (unpublished), August 2006.
[77] S. Lawrence, Free online availability substantially increases a paper's impact, *Nature*, 2001, 411, 521
[78] S. Harnad *et al.*, Comparing the impact of open access versus non open access articles in the same journals, *D-Lib magazine*, 2004, 10 (6), http://www.dlib.org/dlib/june04/harnad/06harnad.html
[79] C. Hajjem, S. Harnad and Y. Gringras, Ten-Year cross-disciplinary comparison on the growth of Open Access and how it increases research citation impact, *IEEE Data engineering Bulletin*, 2005, 28 (4), 39-47
[80] K. Antelman, do open access articles have greater research impact?, *College and research Libraries,* 2005, 65 (1), 372-382
[81] G. Eysenbach, citation advantage of open access articles, *PLoS Biology*, 2006, http://biology.plosjournals.org/perlserv/?request=get-document&doi=10.1371/journal.pbio.0040157
[82] Eysenbach challenges the notion that OA "green" articles (i.e., those in repositories) are more effective than OA "gold" (i.e., those published in OA journals, such as those produced by Public Library of Science) in obtaining high citation counts. It is this part of his paper that produced a furious response from Harnad, much of it focused on particular details. Both authors believe that OA produces a citation advantage, but Eysenbach has presented evidence that casts doubt on Harnad's notion that the "green" route is the preferred route to getting that increased impact.
[83] Oxford Journals, *Assessing the impact of Open Access*, Oxford University Press, 2006.

and it should be noted that there is also clear evidence that numbers of downloads early on is correlated with numbers of subsequent citations[84].

However, despite the intuitive attractiveness of the hypothesis that OA will lead to increased citations because of easier availability, the one systematic study of the reasons for the increased citations – by Kurtz[85] (Source 4.12) – showed that in the field of astronomy at least, the primary reason was not that the materials were free, or that they appeared more rapidly, but that authors put their *best* work into OA format, and this was the reason for increased citation counts.

What impact do citations and IFs have on research funding?

No rigorous study has been published, although there is some limited evidence that funding agencies are sometimes influenced by citation counts. Debackere and Glanzel[86] (Source 4.5) describe the use of citation counts from Web of Science to help the Belgian Government distribute research funds to Flemish universities. This may be the first article to confirm that explicit Government funding decisions have been based on citation analysis. The paper does not explore in any detail whether this policy was controversial and what the political ramifications have been, though it does note that the approach has resulted in changes to funding allocations within the Flemish Universities.

There are also some limited data that suggest that citation counts are also influenced by research funding sources. Bornmann and Daniel[87] (Source 4.3) note that citation counts are highly correlated with receipt of awards from the Boehringer Ingelheim Fonds (BIF), a German foundation for the promotion of basic research in biomedicine. There is not necessarily any suggestion that the BIF funders are influenced by citation counts, but simply that those people who receive BIF funding have higher citation counts than those who applied and failed to receive it. This is significant as it is one of the few documents to consider the link between the two.

All researchers at the University of Oslo must report their published papers in FRIDA (ForskningsResultater, Informasjon og Dokumentasjon av vitenskapelige Aktiviteter)[88]. FRIDA is used to allocate resources to the various departments at the University of Oslo, according to the number of published scientific papers, based on journal impact factors. This appears to be the only use of FRIDA to allocate funds, and is an internal University of Oslo matter.

[84] S. Hitchcock, The effect of Open access downloads on citation impact: a bibliography of studies, http://opcit.eprints.org/oacitation-bibio.html; T. Brody and S. Harnad, Earlier Web usage statistics as predictors of later citation impact, http://eprints.ecs.soton.ac.uk/10713/01/timcorr.htm; T. Brody, unpublished PhD thesis, Southampton University (2006)

[85] M.J. Kurtz *et al.*, The effect of use and access on citations, *Information Processing and Management*, 2005, **41,** 1395-1402

[86] The only article explicitly confirming this approach is K. Debackere and W. Glanzel, Using a bibliometric approach to support research policy-making, *Scientometrics*, 2004, **59** (2), 253-276. This confirms that the Belgian Government uses citation counting as part of its decision-making process for awarding funding to Flemish Universities.

[87] L. Bornmann and H. Daniel, Selection of research fellowship recipients by committee peer review, *Scientometrics*, 2005, **63** (2), 297-320.

[88] http://www.sciecom.org/sciecominfo/artiklar/jakobsson_04_4.shtml

How are IFs affected by the increasing trend to collaborative and multi-location research?

There has been virtually no research in this area, and what research there has been fails to show a convincing cause-and-effect relationship. There is some scattered, but clear evidence, that citation counts to collaborative and multi-location research outputs are higher than to non-collaborative research. Such increases in citation counts that occur appear to have more to do with self-citation rates than anything else. It is well known that authors self-cite to a greater or lesser extent. The more authors a paper has, the more likely it is that there will be a relatively large number of self-citations present, and it is further likely that these citations will be to earlier multi-authored papers. Thus it seems that there is a statistical inevitability about multi-authored papers having a lot of self-citations. Vogel[89] (Source 4.17) is one of the few papers to consider collaboration and IF. The study was of 598 papers on physics published between 1987 and 1994, with at least one author presenting Chilean affiliation. It was found that international collaboration played an important role in the output and that the average IF of journals in which the articles appeared was relatively high and rather constant throughout the period. However, no correlation between the two matters was evaluated. Van Raan[90] (Source 4.16) identifies a link between collaboration and IF, but it is not clear why this occurs. The author's central point is that proper impact measurement must involve corrections for self-citations.

Various bibliometric studies report that multi-institutional or multinational authored papers are more frequently cited than papers that come from a single institute. However, Herbertz's research[91] (Source 4.10) suggests that the conclusion that there is a systematic improvement of scientific success by cooperation is misleading. In a citation analysis of 13 well-known research institutes in molecular biology there was no difference in the average citations per paper with regard to co-operation. In a sub-sample of seven German institutes, the difference found could be explained by self-citations. In another case, all articles of a two-year sample of articles in *EMBO-Journal*, the same phenomenon was observed: differences in the average citations per article with regard to co-operation could be explained by self-citations. Although this paper focuses on citation counts rather than IFs, it is clear that similar results may occur with IF calculations.

Analysis/data validation

At present, the evidence is too sparse, or too scattered to compare or cross-validate in a systematic manner. However, below we have discussed points of consensus and divergence in the existing research base and have drawn conclusions when ppropriate.

[89] See E.E. Vogel, Impact Factor and international collaboration in Chilean physics, 1987-1994, *Scientometrics*, 1997, 38 (2), 253-263; H. Herbertz, Does it pay to cooperate? A bibliometric case study in molecular biology, *Scientometrics,* 1995, 33 91), 117-122.
[90] Van Raan, A. F. J. The influence of international collaboration on the impact of research results: Some simple mathematical considerations concerning the role of self-citations, *Scientometrics*, 1998, 42 (3), 423-428.
[91] Herbertz, H. Does it pay to cooperate? A bibliometric case study in molecular biology, *Scientometrics*, 1995, 33 (1), 117-122

Are traditional journals more likely to be cited than OA journals?

Although there is some evidence suggesting that at the article level, OA articles have greater citation counts than their equivalent subscription-based counterparts[92] (Eysenbach is also important early evidence here), there are methodological factors to be taken into account when extending this argument to consider the impact on journals per se. Some studies indicate by implication that an OA journal in a given subject area will have a greater citation count (or, in some cases, a higher IF, i.e. the average number of citations per article published) than an equivalent subscription-based journal. On the face of it, this seems intuitively obvious. One might expect articles that are readily available free of charge would be likely to receive more citations than articles for which one needs to pay to obtain access.

However, there are serious methodological issues at play; one is not comparing like with like; any two journals will by definition have different IFs, as they have a different selection of articles in them, and the mere fact that one is OA and the other toll-access may not therefore explain any differences in the IF. Furthermore, because some OA journals are new, and IF calculations are based on the last two years' publications in the journal, IFs on newer OA journals can be hard to calculate. What would be particularly helpful would be research based on a journal that had been subscription-based and then switched to OA, or *vice versa*, and to assess the changes in IF that resulted. The only published research in this area appears to be found in a recent study on Oxford Journals[93], and this considered simple citation counts rather than IFs. It is worth noting that *Journal of High Energy Physics*, which started life as OA and then became subscription-based, has continued to show increasing IF over time.

Much of the published research in this area is focused on citation counts of individual articles self-archived in repositories *versus* citation counts of the same or similar articles in a subscription journal. There is some consistency in results showing that OA citations are higher, although there have been a few contradictory (and counter-intuitive) results[94]. Overall, the deposit of research in open access repositories has generally led to a larger number of citations, and earlier citations for those items; however, there are many other factors that influence citation rates, including the reputation of the author, the subject-matter of the publication, the self-citation rate, and, of course, how important or influential the publication is in its own right. Furthermore, the methods adopted by researchers have been subject to criticism[95] because of the mathematical methods employed. Since these criticisms have not been published in detail (third party reports have appeared), it is difficult to judge their validity.

---

[92] S. Lawrence, Free online availability substantially increases a paper's impact, *Nature*, 2001, 411, 521; S. Harnad *et al.*, Comparing the impact of open access versus non open access articles in the same journals, *D-Lib magazine*, 2004, 10 (6), http://www.dlib.org/dlib/june04/harnad/06harnad.html; K. Antelman, do open access articles have greater research impact?, *College and research Libraries,* 2005, 65 (1), 372-382; C. Hajjem, S. Harnad and Y. Gringras, Ten-Year cross-disciplinary comparison on the growth of Open Access and how it increases research citation impact, *IEEE Data engineering Bulletin*, 2005, 28 (4), 39-47; G. Eysenbach, citation advantage of open access articles, *PLoS Biology*, 2006, http://biology.plosjournals.org/perlserv/?request=get-document&doi=10.1371/journal.pbio.0040157; M.J. Kurtz *et al.*, The effect of use and access on citations, *Information Processing and Management*, 2005, 41, 1395-1402; Oxford Journals, *Assessing the impact of Open Access*, Oxford University Press, 2006.

[93] Oxford Journals, *Assessing the impact of Open Access*, Oxford University Press, 2006.

[94] See, for example, C. Creaser, Evaluation of the open access journal experiment: stage 2 report, 2006, in Oxford Journals, *Assessing the impact of Open Access*, Oxford University Press, 2006.

[95] H. Moed, unpublished report (2006)

In summary, the results show that for OA journals, there is (as yet) no consistent evidence of an advantage for being OA rather than being subscription-based, Eysenbach's study of a hybrid journal (PNAS) offering optional OA does indicate higher citation counts for OA articles over the toll-access articles, but further research across disciplines and across a wider range of journals is required to cross-validate this study. For self-archived articles, there does seem to be evidence of a citation advantage over articles in subscription journals, but the main reason for this appears to be the quality of the article itself, rather than anything to do with the open accessibility of the article. Although this is a popular area for research, there is a lack of objective studies of sufficient breadth carried out in a methodologically robust manner. There is a lot of on-going research and this is an area in which new relevant publications appear weekly.

What impact do citations and IFs have on research funding?

This question has not been subject to much literature. Anecdotal evidence[96] suggests that some funding agencies use citation counts as part of the assessment procedure when allocating research funds, e.g. the Wellcome Foundation is believed to use citation counts of individual researchers as part of its procedure when deciding to whom to award funds. The number of publications confirming this approach is tiny, and such evidence as there is comes from policy statements from funding agencies. There is some literature on the citation performance of publications emanating from funding decisions.

Mention should be made of the Research Assessment Exercise (RAE). Officially, the RAE panels evaluating research quality (and thereby influencing subsequent funding allocations) take no account of citation counts and merely assess output on the basis of its inherent quality. In practice, no doubt many panel members do carry out citation counts of items they have to review and the results will no doubt influence their recommendations. There is also a strong correlation between RAE scores and citation counts in a large variety of subject areas, ranging from pure science to the humanities[97]. With future RAEs after the 2008 one likely to be heavily based on metrics, there can be little doubt that in many subject areas, citation counts will feature strongly as a determinant of research funding.

Overall, then, there is very little literature on this topic; more is available on the reverse (impact of funding or source of funds on citation counts). Much more research is needed before any conclusions can be drawn. It is probably safest to conclude that the two are correlated, but that only rarely are citation counts a key factor in funding decision-making. Indeed, it is likely that any major shift in policy to make citation counts a major factor in decision-making on funding would be controversial, just as the use of citation counts to evaluate individuals is controversial.

How are IFs affected by the increasing trend to collaborative and multi-location research?

This question is even harder to answer. Collaborative and multi-location research should in principle have no influence on IF, but multidisciplinary research may well have. This is because, apart from a few leading multidisciplinary journals such as *Nature* and *Science,* most multidisciplinary journals have a relatively low

[96] G. Lewison, personal communication to C. Oppenheim
[97] See M. Norris and C. Oppenheim, Citation counts and the Research Assessment Exercise V: Archaeology and the 2001 RAE, *Journal of Documentation*, 2003, 59 (6), 709 – 730, and references cited therein.

IF compared with single discipline journals. The few articles that have appeared on this topic are focused on particular subject areas and/or geographic regions[98].

Overall, little research has been carried out in this field, but there is some evidence of a link, although self-citations appear to be the primary cause. In other words, when involved in collaborative research, a particularly large number of authors cite themselves, and that in turn improves citation counts and/or IFs. There is no suggestion of a causal link between self-citing and collaboration, and therefore much more research is needed in this area.

Gaps in the data

Although quite a lot of evidence has been collected regarding the effect of OA (whether in the form of OA journals or as self-archived articles) on citation counts, much of it is scattered, uses inconsistent methods and covers different subject areas, although not surprisingly high-energy physics has been one major area of research. The research rarely takes into account other factors that can affect citation counts, and it is difficult to see how these factors can be systematically evaluated. A rare exception is Kurtz *et al.* (Source 4.12), who used sophisticated statistical methods to identify the main factor leading to increased citations in self-archived astronomy articles.

The whole area of the relationship between citation counts and scholarly communication channels is confused because of problems associated with quality bias (e.g., if scholars tend to self-archive only their best work, as suggested by Kurtz *et al.*; alternatively, it may be that only the best journals are OA). In other words, differences in citation counts and IFs may simply reflect the quality of the materials under study rather than having anything to do with the channel by which the material is made available.

An important new metric, the *h-index* is extremely useful for evaluating individuals' impact[99], but no studies to date have considered whether a researcher's *h-index* score is affected by the number of OA articles he/she publishes. The *h-index* is easily calculated; a score of 1 means the author has one article that has been cited at least once; a score of 2 means the author has at least two articles that each have been cited at least two times, and so on. Thus an *h-index* score of 20 means that author has published 20 articles, each of which has ben cited at least 20 times.

Overall, we concur with Craig's introduction that "the problems with measuring and quantifying an Open Access advantage are significant. Articles….cannot be OA and non-OA at the same time. Further, the variation of citation counts between articles can be extremely high, so making controlled comparisons of OA vs. non-OA articles nigh on impossible" and with his conclusion that "as any Open Access advantage appears to be partly dependant on self-selection, the more articles that are {self-}archived…so you'd expect to see any Open Access advantage reduce. Authors self-archiving in the expectant belief that each and every paper they archive will receive an Open Access advantage of several hundred percent are going to be sorely disappointed."

---

[98] See E.E. Vogel, Impact Factor and international collaboration in Chilean physics, 1987-1994, *Scientometrics*, 1997, 38 (2), 253-263; H. Herbertz, Does it pay to cooperate? A bibliometric case study in molecular biology, *Scientometrics,* 1995, 33 91), 117-122.

[99] B. Cronin and L. Meho, Using the *h*-index to rank influential information scientists, *Journal of the American Society for Information science and Technology*, 2006, 57 (9), 1275-1278, and references cited therein

Finally, it is worth noting that all researchers in the field are agreed that if the vast majority of scholarly publications become available in OA form, no citation advantage to OA will be measurable. Thus, what OA advantage there is will prove to be temporary if OA does become the standard mode of publication.

The other topic areas have hardly been studied at all, and there is a need for further systematic research on, for example, citation counts and IFs of journals that were toll-access and went OA[100] (or *vice versa*), or of articles that are available in both forms *versus* articles that are only available in one of the forms.

Little or no research has been carried out on collaboration and its effect on citations, or on funders' approaches to citations as a metric for decisions.

It is clear, therefore, that much research is needed in these areas before clear conclusions can be drawn.

Sources evaluated in detail in Appendix 1

Source 1: Antelman, K. Do open access articles have greater research impact? *College and Research Libraries*, 2005, 65 (1), 372-382

Source 2: Bauer, K and Bakkalbasi, N, An examination of citation counts in a new scholarly communication environment, *D-Lib magazine*, 2005, 11(9), http://www.dlib.org/dlib/september05/bauer/09bauer.html

Source 3: Bornmann, L and Daniel, H. Selection of research fellowship recipients by committee peer review, *Scientometrics*, 2005, 63 (2), 297-320

Source 4: Brody, T, Harnad, S. Earlier Web Usage Statistics as Predictors of Later Citation Impact. http://eprints.ecs.soton.ac.uk/10713/01/timcorr.htm

Source 5: Debackere, K and Glanzel, W. Using a bibliometric approach to support research policy-making, *Scientometrics*, 2004, 59 (2), 253-276.

Source 6: Dhawan, SM and Gupta, BM. Evaluation of Indian physics research on journal impact factor, *DESIDOC Bulletin of Information Technology*, 2005, 25 (3), 3-7.

Source 7: Eysenbach, G. Citation advantage of open access articles, http://biology.plosjournals.org/perlserv/?request=get-document&doi=10.1371/journal.pbio.0040157

Source 8: Hajjem, C., Harnad, S. and Gingras, Y. (2005) Ten-Year Cross-Disciplinary Comparison of the Growth of Open Access and How it Increases Research Citation Impact. *IEEE Data Engineering Bulletin* 28(4) pp. 39-47. http://eprints.ecs.soton.ac.uk/11688/

Source 9: Harnad, S, Brody, T, Oppenheim, C *et al.*, Comparing the impact of open access versus non open access articles in the same journals, *D-Lib Magazine*, 10,(6), 2004, http://www.dlib.org/dlib/june04/harnad/06harnad.html

Source 10: Herbertz, H. Does it pay to cooperate? A bibliometric case study in molecular biology, *Scientometrics*, 1995, 33 (1), 117-122

Source 11: Hitchcock, S, The effect of open access downloads on citation impact: a bibliography of studies, Opcit project, 2004 – http://opcit.eprints.org/oacitation-biblio.html

[100] The important Oxford Journals research needs to be followed up in a few years' time to see whether the citation count changes noted were maintained, and similar studies need to be carried out with other publishers, such as Elsevier and Springer, that have adopted OA or Open Choice approaches to some or all of their journals.

Source 12: M.J. Kurtz *et al.*, The effect of use and access on citations, *Information Processing and Management*, 2005, 41, 1395-1402

Source 13: Lawrence S., Free online availability substantially increases a paper's impact, *Nature* 411, 521 (31 May 2001)

Source 14: Oxford Journals (2006). *Assessing the impact of Open Access*, Oxford University Press.

Source 15: Thomson Scientific. *The impact of open access journals – a citation study from Thomson ISI*, 2004: http://scientific.thomson.com/ts/media/presentrep/acropdf/impact-oa-journals.pdf

Source 16: Van Raan, A. F. J. The influence of international collaboration on the impact of research results: Some simple mathematical considerations concerning the role of self-citations, *Scientometrics*, 1998, 42 (3), 423-428.

Source 17: Vogel, E. Impact factor and international collaboration in Chilean physics : 1987-1994 , *Scientometrics*, 1997,. 38 (2), 253-263

Area 5: Disciplinary differences

Executive summary

Key questions

This section reviews the evidence in two areas:

1. Is there any difference in the way researchers active in different disciplines use journals as readers?
2. Is there any difference in the publishing habits of different disciplines?

Findings in Area 5 are closely related to issues in Area 3 of this report and there is a high degree of overlap in the major sources consulted.

The evidence

Surveys of authors and readers (again mostly in higher education environments rather than industry) provide the richest data on these issues, but the focus, population samples, and methods of individual studies, together with the form and content of survey questions, obviously differ so direct comparison of results can only be tentative.

The survey evidence suggests that journal articles are most important in the sciences and social sciences, but that books are more important in the arts and humanities. There is some evidence of a trend towards greater convergence in the use of research resources between disciplines, but this finding needs further validation.

All researchers appear to have similar levels of access to the journal materials they need. The issue of ease of access to journals shows little meaningful variation by discipline – around 50% of all researchers, regardless of discipline, experience problems.

Faced with such barriers, only around a quarter of researchers encountering difficulties (both arts and sciences) never try to obtain articles via another method. Inter-library loan is the most popular alternative route in both the arts and the sciences. Those in the sciences are more likely to also turn directly to article authors for help than their counterparts in the arts.

All authors, irrespective of discipline, claim that career advancement and peer-to-peer communication are the most important reasons for publishing.

There is a clear consensus across sources on which factors influence an author's choice of journal in which to publish; impact factor, reputation and peer review were reported to be of primary importance to all authors, again irrespective of discipline.

Publication delays were identified as an obstacle impeding the publishing process in 1999 by nearly 50% of researchers in all disciplines, but most expected or hoped that electronic publishing with rapid peer review would remove this hurdle. We have no up-to-date evidence to validate whether this has in fact proved to be the case. Researchers in all disciplines anticipate that electronic dissemination of research will be increasingly important through to 2015.

The gaps

As in Area 3, analysis of transaction log data, recording usage activity (segmented by discipline), would add significant value and rigour to the information available in this area. The survey data we do have access to focuses quite narrowly on attitudes and preferences in relation to journals rather than exploring how journals are actually used in the context of researchers' professional lives.

The evidence we are able to draw on points to disciplinary convergence rather than divergence in researcher usage of and attitudes towards journals, whether from a reader or author perspective.

As highlighted in Area 3, a more large-scale exploration of the nature of discipline-specific journal usage – and most importantly the evolving role and value of journal articles alongside alternative modes of scholarly communication - would bring more depth to an area in which our understanding is superficial.

Introduction

This Area focuses on disciplinary differences in the use of journal articles and should be read in conjunction with Area 3 'Usage'. The sources reviewed for Areas 3 and 5 overlap to a very considerable extent and are therefore presented in a single section in Appendix 1 of this report.

Scope and definitions

This area of the project investigates two areas:

- Is there any difference in the way researchers active in different disciplines use journals as readers?
- Is there any difference in the publishing habits of different disciplines?

The reviewers intended to use fairly broad discipline categories in this analysis, namely Humanities, Social Sciences and Sciences. However, different sources use different definitions, some using only two categories and others more. This is indicated in the discussion.

Use of journal *articles* is discussed when breakdown to this level is possible. However, some analyses only consider journal use in general. Furthermore, some surveys consider academics' use of resources in general, of which journals are only a part.

As with Area 3, the majority of respondents were academics in higher education institutions and some of the surveys were restricted to this population.

Key sources

All of the sources reviewed for Area 3 were also mined for any discipline-level analysis. The author surveys were most useful; generally, the level of analysis reported in data log analyses was not sufficient to give insight into disciplinary differences. Surveys provide the richest data, but the focus, population samples, and methods of individual studies, together with the form and content of survey

questions, obviously differ so direct comparison of results should be viewed with some caution.

What are the major differences between disciplines in journal use?

Most of the sources reviewed for Area 3 provide some insight into journal use by different disciplines. However, the analysis section in this report will focus on the following sources which provide the greatest level of break-down by discipline:

Source 9 **Rightscom Ltd (2005). JISC Disciplinary differences report.**
This is a survey of 780 academics in UK HEIs with the aim of informing JISC on the barriers and opportunities facing researchers in relation to their access to research resources. The survey covered all research resources, so journal use is reported in this context. The study uses RAE 'Unit of Assessment' categories for analysis and summary results are reported under 5 groups, but results are also provided at subject level.

Source 13 **Key Perspectives Ltd., (2002). Authors and electronic publishing: the ALPSP research study on authors' and readers' views of electronic research communications.**
This is a web-based survey of 14,643 authors. Authors were solicited by approaching publishers for details of contributors to a selection of their journals. In all 23 publishers participated, including society, university, and commercial players. The survey response rate was 8.5 per cent (1,246 responses). Findings include the views of respondents in their role as readers of journals as well as in their role as authors/contributors. All responses are presented by discipline (Arts and Sciences).

Other sources reviewed provide only limited break-downs of data by discipline (See List A). Evaluation sheets for all sources are provided in Appendix 1.

What are the major differences between disciplines in publishing preferences and habits?

Three sources provide valuable data on the views of authors, with break-down by discipline:

Source 5 **Key Perspectives Ltd., (1999). What do authors want: the ALPSP research study on the motivations and concerns of contributors to learned journals.**
This is the only source which presents the views of academics as authors only. It also focuses on print journals, although the issue of electronic publishing is investigated.

Source 9 **Rightscom Ltd (2005), JISC Disciplinary differences report.**
Survey of academics in UK Higher Education Institutions to inform JISC on the barriers to publication of research.

Source 13 **Swan, A and Brown, S. (2002), Authors and electronic publishing: the ALPSP research study on authors' and readers' views of electronic research communications.**
Follow up to the 1999 survey focusing on electronic publishing. Not all of the issues raised in the 1999 survey are included.

Analysis/data validation

What are the major differences between disciplines in journal use?

The Rightscom study (Source 3/5.9) tentatively concludes that there is some convergence in the use of research resources between disciplines, possibly due to increased availability and ease of use of electronic publications. However, the survey did find that journal articles were most important in the sciences and social sciences, but that books were more important in arts.

There seemed to be little difference between disciplines in ease of access to journals. All groups showed a near 50:50 split between those who did encounter problems and those who did not (see Table 5.1) The major problem reported was access to journal articles because 'the library does not subscribe to required titles'. This was reported by a larger proportion of scientists, but should be considered in the context that journals are the single most essential resource in sciences.

Table 5.1. Do you encounter problems gaining access to the resources you need to carry out your research?

	Yes/%	No/%
Medical and biological sciences	53.5	47.5
Physical sciences and engineering	42.4	57.6
Social sciences	46.7	53.3
Languages and area studies	48.0	52.0
Arts and humanities	53.4	46.6

Swan and Brown (Source 3/5.13) cover access slightly differently. As the focus of the survey was the use of electronic journals, issues of remote access are investigated. Both disciplines consider the ability to access from the desktop as very important (76.3 per cent of Science respondents and 71.4 per cent of Arts respondents). However, there are differences in views regarding the location of additional access: 52.8 per cent of Arts respondents considered ability to access electronic journals from home 'very important' compared with 37.4 per cent of Science respondents; and 62 per cent of Arts respondents considered ability to access at any time 'very important' compared with 54.2 per cent of Science respondents.

Unfortunately, satisfaction with access is not reported directly. However, respondents were asked how they obtain articles 'not available to them' and these results indicate researchers' motivation for obtaining articles. The results are summarised in Table 5.2. Science respondents are slightly more reluctant to order from publishers or commercial document supply outlets than Arts respondents, but are much more likely to request articles from authors. Although more Arts respondents 'very often' do not try to obtain articles, the percentage of people who never try is the same for both disciplines. The proportion of researchers who do try to obtain articles is therefore quite high, and suggests a high regard for the value of journal articles, as well as awareness of alternative routes to access them, inter-library loan being the most widely used option.

Table 5.2. How do you obtain articles not available to you?

	Arts		Sciences	
	Very often (%)	Never (%)	Very often (%)	Never (%)
Order from publisher/commercial document supply	~4.0	~50.0	~2.0	65.0
Inter-library loan	42.2	7.4	35.7	9.7
Request from author	5.0	28.3	16.2	14.2
Don't try to obtain	9.2	26	3.9	26.5

Library provision is raised under 'perceived problems' with electronic journals. Uncertainty about continuing access was a major concern for both arts and science respondents, with very little difference between disciplines: 36.2 per cent of science respondents and 34.6 per cent of arts respondents were 'very concerned' about this potential barrier to access.

Table 5.2 indicates that researchers have a tendency still to obtain articles they need, even if they are not immediately available. The Rightscom (Source 3/5.9) study addressed the importance of journal articles more directly by asking researchers 'what is the single most essential resource you use' and 'what is the most heavily used resource'. Similar results were received for both. Journal articles were most important in the sciences and social sciences, but books were more important in arts.

What are the major differences between disciplines in publishing preferences and habits?

Key Perspectives (Source 3/5.5) found the two most important reasons for publishing were 'Communication with Peers' and 'Career Advancement'. There was little difference between disciplines. Although the Rightscom Survey (Source 3/5.9) did not directly address this question, the survey did find that journal articles were an equally important method of dissemination for both peer-to-peer communication and RAE score: other methods were not considered equally influential for both purposes.

There was consensus across different sources regarding the factors which influence authors in choosing a journal. Factors such as impact factor, reputation and peer review were reported to be of primary importance to all authors, irrespective of discipline.

Key Perspectives (Source 3/5.5) found publication delays are seen as a major obstacle by both Science and Arts respondents (about 45 per cent in each discipline noted this as a 'major obstacle'). Peer review was also noted as a major obstacle by just over 15 per cent in each discipline respectively, but this was also related to the basic issue of delays. Publication delays were also noted as an obstacle in the Rightscom survey, although 'rejection rates for highly rated journals' also scored highly as a barrier to getting published.

There was a high level of agreement on the future of scholarly publishing, both between disciplines and over the 6 years covered by the different surveys.

Key Perspectives (1999) found support for the current system of publishing, particularly from scientists: 71.6 per cent of Science respondents and 60.4 per cent of Arts respondents thought publishing *should* continue in its current form. There was little support for elimination of the peer review process from either discipline, although both favoured 'electronic publishing with rapid peer review':

about 50 per cent of all respondents thought this would happen, and about 38 per cent 'would like' it to happen – with slightly more support from science respondents than other disciplines.

Although later studies worded questions about the future of publishing differently, there was most support for 'traditional print plus electronic' from both disciplines in the 2002 survey (Source 3/5.13). The Rightscom survey (Source 3/5.9) also found support for the journal article (although method of delivery was not specified) – respondents from both science and arts disciplines thought the journal article would remain relevant during the next 10 years. This survey also noted support for electronic dissemination with about 75 per cent of respondents across all disciplines agreeing that this would be increasingly important during the next 10 years.

Gaps in the data

What are the major differences between disciplines in journal use?

As with Area 3, there is a lack of published statistics on journal use at the article level, although such data are collected in transaction log analyses collected by publishers and other research organisations. It would be possible to analyse such data at a discipline level to help gain additional insight into actual researcher behaviour to cross-validate with market surveys. It may be argued that for both Area 3 and Area 5 there is a need for more large-scale exploration of usage (including the nature of such usage) that provides consistent and comparable quantitative and qualitative data across disciplines.

What are the major differences between disciplines in publishing preferences and habits?

This area is well covered by the surveys carried out by Key Perspectives for ALPSP in 1999 and 2002, and by Rightscom for JISC in 2005. However it is important to keep information up to date.

Sources evaluated in detail in Appendix 1

Source 2	Education for Change, SIRU, and Research Partnership, (2002). Researchers use of libraries and other information sources, 2002
Source 5	Key Perspectives (1999), What authors want: the ALPSP research study on the motivations and concerns of contributors to learned journals
	Source 9 Rightscom Ltd., (2005). JISC Disciplinary differences report
Source 12	Rowlands, I and Olivieri, R., (2006). Journals and scientific productivity.
Source 13	Key Perspectives Ltd., (2002). Authors and electronic publishing: the ALPSP research study on authors' and readers' views of electronic research communications

Area 6: Cost and impact of alternative formal dissemination models

Executive summary

Key questions

The key questions to be addressed in Area 6 are specified as:

1. What are the costs involved in publishing open access journals? To what extent are these different from those of publishing conventional journals?
2. What is the impact of digital repositories, institutional or thematic, on the economics of journal publishing?

The evidence

Sources in this area are diverse in focus, and more often than not provide glimpses of only limited areas of the topics under examination (as in, for example, close scrutiny of a small handful of journal titles, or examination of a narrow range of disciplines). The paucity of sources means that establishing evidence-based causal relationships in key areas cannot currently be done and, similarly, extrapolation from restricted samples to wider communities is currently not possible.

On the issue of costs, any evaluation of the impact of alternative models presupposes an understanding of existing models. However, the evidence about the costs of the traditional journal publishing process (described in Area 2 of this study) does not provide solid comparators against which OA costs can be set.

The focus of the current debate about the relative costs of an OA publishing system versus publishing conventional journals seems to be gradually changing. Our evidence review suggests a degree of acceptance that many of the components of cost are common to both principal models and can therefore be cancelled out in the 'equation' – for example, 'first copy' costs at the beginning of the cycle, and server and software costs at the other end.

The research base around 'costs' appears to be getting broader in scope. Work has now been undertaken on how new OA journal publishing models such as 'author pays', will be funded – and by whom. There is evidence suggesting that a straightforward institutionally-based solution would potentially be inequitable, concentrating the cost on a relatively small number of research-intensive institutions.

In 2005, the average number of items held in repositories was estimated to be 'a few hundred' (with the exception of the Netherlands and possibly the USA, for which details were not available). These items tend to be very diverse in nature, including to a very considerable extent teaching material as well as research contributions. An obvious exception is the long-established thematic repository, arXiv in the field of high-energy physics and related areas.

Whilst some evidence does suggest that these repositories are an important new factor in the journal cancellation decision process, and one which is growing in significance, there is no research reporting actual or even intended journal subscription cancellation as a consequence of the growth of OA self-archived repositories.

Subscriptions are reported to have been declining over a period of 10+ years, but for a number of reasons. Proving or disproving a link between availability in self-archived repositories and cancellations will be difficult without long and rigorous research. In this connection, the outcome of research recently announced by the Research Councils UK (RCUK),[101] with the co-operation of Macmillan, Blackwell and Elsevier, will be eagerly awaited, even though a report is not due until late 2008.

The gaps

On the issue of costs, the critical gap will be for a wide-ranging study on the funding and money flow implications of new publishing paradigms, particularly if the author-pays model becomes well established.

There is no evidence as yet to demonstrate any relationship (or lack of relationship) between subscription cancellations and repositories. Work in this field would need sufficient, representative and balanced samples, and the collaboration of all stakeholders, including especially research institutions and publishers. Any such study will need to be maintained over a fairly extended period, with regular reports, since it seems likely that the position could change with time if the contents of self-archiving repositories become progressively more comprehensive.

Similarly, more carefully conceived work on the impact of both OA journals and self-archiving on the quality of research communications, especially on the peer review system, will be required.

Introduction and scope

The specification for this study elaborates on Area 6 of the work as follows:

- What are the costs involved in publishing open access journals? To what extent are these different from those of publishing conventional journals?
- What is the impact of digital repositories, institutional or thematic, on the economics of journal publishing?

Two points need to be made immediately:

- Firstly, any evaluation of the impact of alternative models presupposes an understanding of existing models, and, for any economic assessment to be made, that understanding must extend to the cost profiles of existing models. This latter topic is discussed in Area 2 of this study, where the wide diversity of print publishing costs is examined in depth. For a full discussion of evidence relating to costs under the 'traditional' publishing model, therefore, reference should be made back to Area 2.
- Secondly, while there is considerable overlap between the issues raised by open access and digital repositories, it is important to remember the distinction (see ***Definitions***, below), and to consider them separately, when necessary, as well as together.

This review of 'alternative formal dissemination models' seeks to identify and evaluate evidence relating both to open access journals and to self-archiving

[101] http://www.rcuk.ac.uk/access/2006statement.pdf

digital repositories. It has to be recognized, however, that these are live and evolving phenomena, so relevant new data may become available at any time.

As stated elsewhere in this report, the costs and issues relating to the long-term preservation of journals (say 10+ years), as opposed to storage for access (including the availability of backfiles), are considered to be outside the scope of this study.

Definitions

For the purpose of this report, the following definitions have been adopted:

- **Alternative formal dissemination models.** It is clear from the terms of reference, that for the purpose of this study this expression refers specifically to open access journals and self-archiving digital repositories (see below). Self-evidently (but for the avoidance of doubt), this rules out consideration of the rôle of informal types of scholarly communication.

- **Open access.** The Budapest Open Access Initiative (BOAI), which resulted from a meeting convened by the Open Society Initiative[102] (OSI) in December 2001 and which was promulgated in February 2002, provides a widely acknowledged definition of OA, *viz*:

  > *By 'open access' to this literature, we mean its free availability on the public internet, permitting any users to read, download, copy, distribute, print, search, or link to the full texts of these articles, crawl them for indexing, pass them as data to software, or use them for any other lawful purpose, without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself.*[103]

  Clearly, OA is in direct contrast to the 'traditional' publishing model, under which published material is supplied only to users who pay for access (whether personally or via the organization to which they are affiliated).

  Two routes to OA can be differentiated – **OA publication** within journals (possibly subject to conditions), and **OA self-archiving** (in institutional or other repositories). It should, however, be borne in mind, that within these two basic divisions of OA, variants exist. In fact, Willinsky[104] has identified nine different sub-species of pure electronic or hybrid print-plus-electronic OA, although the scope of his OA universe, which includes e-prints is rather wider than that of some other experts.

  Most of the literature reviewed in this Area of this study relates to OA publication. Indeed, the literature reviewed often applies mostly to issues surrounding the so-called 'author-pays' model of OA publication. Where literature concerns itself with OA self-archiving, this focus is indicated, the most obvious instances being in the section below dealing specifically with repositories (Section 6.3.3).

---

[102] http://www.soros.org/about
[103] http://www.soros.org/openaccess/read.shtml
[104] Willinsky, J. The Nine Flavours of Open Access Scholarly Publishing. *J Postgrad Med*, 2003;49:263-267: http://www.jpgmonline.com/article.asp?issn=0022-3859;year=2003;volume=49;issue=3;spage=263;epage=267;aulast=Willinsky

- **Digital repository.** This term refers to a location where information resources (whether born-digital or digitized from previous print versions) are placed for the purpose of access by users. Some repositories, but not necessarily all, also concern themselves with longer-term preservation. The location could be a single server or a series of linked servers in one or more physical places. Repositories may hold content falling into any of a range of categories, including personal, institutional (e.g. college, university, museum), national, or thematic (e.g. high-energy physics, classical literature). Both Lynch[105] and Ware[106] (Source no. 6.9) provide useful definitions of institutional repositories which are cited by others.

Key sources

Studies based upon solid up-to-date primary data are in short supply in Area 6 of this work. Much has been written in polemic vein concerning open access and digital repositories, both separately and together. An attempt was, therefore, made for the purpose of this study to seek out the basis of fact and evidence upon which observers and advocates – whether of radical change or of the *status quo* – have constructed their arguments. In consequence, more items were examined and put on one side than were evaluated in detail (see Appendix 1 of this report for the full evaluations). Much of the material which has been evaluated is of North American origin, and there is little which has a primarily UK focus. Given the international nature of journal publishing, however, most of the sources evaluated have at least some relevance to the UK scholarly communications marketplace.

Costs involved in publishing open access journals

Direct evidence

An important publication, which is relatively recent, which is based upon a survey with an adequate sample, and in which the workings and conclusions are transparent, is the October 2005 Kaufman-Wills report[107] on OA publication (Source 6.3). This work was sponsored by the UK-based Association of Learned and Professional Society Publishers (ALPSP), by the American Association for the Advancement of Science (AAAS), and by HighWire Press/Stanford University Libraries, with co-operation also from the Association of American Medical Colleges (AAMC). Key contributions to the debate which followed publication – from Jan Velterop, Frederick Friend, and BioMed Central – can be found via issue 91 of the *SPARC Open Access Newsletter*[108]. This item also includes the text of an interview by Peter Suber with Cara Kaufman, who addresses points made during post-publication peer review, some of which had led to the publication of an addendum providing additional data and correcting a significant error in one particular area. Taken together, the updated publication and the debate provide valuable data on the present state of OA publishing, including financial aspects,

---

[105] Lynch, C. Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age. *ARL bimonthly report*, 226, February 2003, 1-7. http://www.arl.org/newsltr/226/ir.html

[106] Ware, M. *Pathfinder Research On Web-Based Repositories: Final Report*, January 2004. http://www.palsgroup.org.uk/palsweb/palsweb.nsf/0/8c43ce800a9c67cd80256e370051e88a/$FILE/PALS%20report%20on%20Institutional%20Repositories.pdf

[107] Kaufmann-Wills Group LLC on behalf of ALPSP, AAAS, HighWire Press and AAMC. *The facts about open access: a study of the financial and non-financial effects of alternative business models for scholarly journals* [Including post-publication addendum with additional, corrected data and analysis on peer-review] (October 2005). http://www.alpsp.org/publications/FAOAcompleteREV.pdf

[108] http://www.earlham.edu/~peters/fos/newsletter/11-02-05.htm#kaufman (2 November 2005)

based upon nearly 500 responses to the authors' survey, supported by 22 case studies. The resultant data, though international in scope, are certainly relevant to OA publication in the UK.

Evidence that one high-profile OA publishing pioneer at least – the Public Library of Science (PLoS) – has not been able to get its costs and revenues quickly into an appropriate equilibrium was reported very recently. Figures published in *Nature*[109] in June 2006 showed that the early break-even hoped for by PLoS is some way off, with author fees and advertising revenues covering only 35 per cent of total costs. One consequence was that PLoS was constrained to increase its author fees from US$1,500 per article to up to US$2,500. Even so, Mark Patterson of PLoS has also been reported as saying that the increased author fees still may not be enough to cover costs[110]. This statement is in contrast to the more sanguine expectation expressed by Dr Harold Varmus to the House of Commons Science and Technology Committee on 8 March 2004[111], when he indicated that PLoS would be self-sustaining within about two and an half years or so from that date.

Going back to 2004, Hedlund, Gustafsson and Björk[112] (Source 6.2) published data from a survey of OA journal editors which had elicited 60 responses in May 2003. The analysis of the scientific publication value chain and how it may change under an OA publication model is interesting. Data collected by the survey, however, are largely qualitative, except for the time spent by individuals involved in journal management and editorial processes. There is, however, no attempt to translate these into monetary values, and indeed the authors took the view that 'OA publishing in its current form is presumed to be a voluntary activity, to some extent at least'. Hence, although the time taken up by the processes examined is of interest, the article contributes virtually nothing to an understanding of the cost base of OA publication. For an assessment of the costs of the authorship element of publication, which is common to all models, reference should be made to Area 2 of this study.

Indirect evidence

Although not directly addressing the issue of costs, various other sources worthy of note provide valuable data on background, on usage and on authors' attitudes. Clearly, the last two will be important factors affecting cost profiles and determining the lasting success (or otherwise) of OA publishing.

The most recent contribution to data on OA journals was published in June 2006 by Oxford University Press[113] (Source 4.14, in Area 4). This publication combines three short reports relating to one or more of three OA journals – *Nucleic Acids Research* (*NAR*), *Journal of Experimental Botany* (*JXB*), and *Bioinformatics*. The first study was based on a solid response from 1,144 individuals who had participated in some aspect of the *NAR* editorial process (so, not mere readers). It tested attitudes, and probed funding and usage questions. In the second report, usage and citation patterns were examined for all three journals, confirming that articles published in OA journals generally attract higher levels of

[109] Butler, D. Open-access journal hits rocky times, in *Nature*, 20 June 2006, http://www.nature.com/news/2006/060619/full/441914a.html

[110] *Ibidem*. 'We will continue to rely on philanthropic grant support for the foreseeable future'.

[111] House of Commons Science and Technology Committee. *Scientific Publications: Free for all? Volume II: Oral and Written Evidence.* (HC 399-II). Page 27: Response to Question 188. http://www.publications.parliament.uk/pa/cm200304/cmselect/cmsctech/399/4030808.htm

[112] Hedlund T, Gustafsson T, Björk B. The open access scientific journal: an empirical study, July 2004. http://oacs.shh.fi/publications/199-210.pdf

[113] Oxford Journals, Oxford University Press. *Assessing the impact of open access – preliminary findings from Oxford Journals*, June 2006. http://www.oxfordjournals.org/news/oa_report.pdf

article downloads than subscription-only journals. The quantity of data concerning citations was still, however, considered insufficient to draw reliable conclusions. The last of the reports looked at deep-log data concerning *NAR*, which indicated that opening the site to search engines had generated appreciably more usage than the move to OA.

Possibly the most interesting part of the Oxford experience is not documented in the above-mentioned report, but formed part of the introductory presentation by Martin Richardson at the event hosted by OUP to launch the report[114]. Richardson revealed that income per article for *NAR*, which went fully OA from January 2005, had gone down from $4,647 in 2004 to $3,622 in 2005. It was also mentioned that dealing with payments from individual authors had generated an increased administrative burden (although it has to be recognized that, on the other hand, subscription administration would no longer be applicable). Caution, however, has to be exercised, for while the Oxford experiments are valuable, they still only relate to three journals in the early stages of OA. Nevertheless, taken with the disappointing financial out-turn for PLoS referred to above, the OUP financial experience suggests that viability for OA journal publishing will not come quickly and easily.

Earlier research in this field has largely been superseded because of the speed of developments. Nevertheless, some useful insight into authors' attitudes to OA publication were collected in a study in 2004 by Swan and Brown[115] (Source 6.5) which elicited over 300 responses, though authors' awareness and experience of both OA journals and self-archiving has moved on since.

Extent to which costs are different from those of publishing conventional journals

In April 2004, a report commissioned by the Wellcome Trust from the consultancy SQW Ltd[116] (Source 6.7) argued not only that the OA author-pays model is a viable alternative, but that it appears to be less costly. The report constructed its arguments in favour of OA by assembling data from a number of earlier sources and from additional discussions with individuals in senior positions in STM publishing. To the extent that the report provides a round up of data on 'first copy costs', which are common to all models, it is useful. How much genuinely new data the exercise brought to the arena, however, is not at all clear from the report, since amongst the 12 sources listed in the References at the end of the report, 11 are previously published sources, and only one acknowledges unpublished data from a practising publisher.

Disappointment was expressed in some quarters over a perceived partiality (in both senses – i.e. biased and incomplete) in the SQW report, in that it did not address, *inter alia*, the question of overhead costs, continuing investment in systems, market research and other costs associated with the launch of new journals, and the purpose and application of profits/surpluses for both commercial and non-commercial publishers. Nor did the report properly address the considerable diversity of publishing costs.

---

[114] Slide set: 'Experimenting with open access publishing: overview', presented 5 June 2006.

[115] Swan, Alma and Brown, Sheridan. Authors and open access publishing. *Learned Publishing*, 2004, 17(3), 219-224. http://puck.ingentaconnect.com/vl=19271709/cl=11/nw=1/fm=docpdf/rpsv/cw/alpsp/09531513/v17n3/s7/p219

[116] SQW Ltd for the Wellcome Trust, *Costs and business models in scientific research publishing*, April 2004. http://www.wellcome.ac.uk/assets/wtd003184.pdf

Two subsequent papers which are worthy of note contributed a well-informed discussion of the full spectrum of costs associated with OA publication and 'conventional' publishing (see also Area 2 concerning the latter). One of these papers, by King and Tenopir[117] (Source 2.10, in Area 2), emanated from the academic community the other, by Morris[118], from the publishing community (Source 2.3, in Area 2). A reading of these sources tends to point towards the conclusion that the costs are likely to be relatively little different, with OA publication being perhaps slightly less expensive. Both papers, however, raise questions over sources of funding for author-pays fees, and especially how the changed money-flows in the new model could potentially cause inequitable imbalances to the detriment of research-intensive institutions and to the advantage of organizations which are consumers of research information but which do not generate significant authorship.

The issue of the possible economic effect of OA publication on institutions with high scholarly output was particularly highlighted in an article in *Serials* in March 2006. Dominguez[119] (Source 6.1) examined how applying OA publication costs for six journals to the scholarly output of her institution (CERN) would compare with current subscription costs. She used two cost models, including author-only payment, and concluded that the OA costs would turn out substantially higher than the subscriptions. Whilst it has been suggested that this article may not have compared like with like[120], and is restricted to only one institution and six journals, it illustrates well the potential funding issue.

During the research for the present report, a similar exercise conducted by the Council for the Central Laboratory of the Research Councils (CCLRC) in the UK also came to light. To inform its library strategy, CCLRC undertook an exercise some years ago that estimated the cost of gold (author institution pays) OA publication for the organization. Using the then published costs of relevant journals, their estimate indicated that it would cost CCLRC something like three times as much as existing library subscriptions. This is because CCLRC is a research output productive institution[121]. The complexity of the issue for funding bodies is demonstrated by the fact that the average number of funders per piece of research in the City University Research Outputs Database (ROD)[122], which covers biomedical and healthcare research, comes out at 2.7.

Helpful advice for learned societies in a US context is provided by Willinsky[123] (Source 6.10). The data provided relate to only 20 associations and are now ageing somewhat (taken from tax returns relating to 1999 or 2000). They are also 'top level', in that costs, in particular, are not broken down at all. Nevertheless, the overview of society economics and the focus on cost saving and alternative revenue sources are useful contributions.

---

[117] King, D W, and Tenopir, C: An evidence-based assessment of the 'author pays' model, *Nature* Web Focus, June 2004, http://www.nature.com/nature/focus/accessdebate/26.html

[118] Morris, S. The True Costs of Scholarly Journal Publishing. *Learned Publishing*, 2005, 18(2), 115-126. http://alpsp.publisher.ingentaconnect.com/content/alpsp/lp/2005/00000018/00000002/art00006;jsessionid=2p98rnbw2m1ne.victoria#avail

[119] Dominguez, Magaly. Economics of open access publishing, in *Serials*, 19(1), March 2006. http://serials.uksg.org/(pquu4jmkt5y0kc45qjzblr3k)/app/home/contribution.asp?referrer=parent&backto=issue,14,20;journal,1,34;homemainpublications,2,2;

[120] See comments addressed to the liblicense listserv by Frederick Friend: http://www.library.yale.edu/~llicense/ListArchives/0605/msg00054.html

[121] Interview with Professor John V Wood, Chief Executive of CCLRC, and Professor Keith Jeffery, Director of IT and International Strategy, on 8 August 2006.

[122] http://www.soi.city.ac.uk/organisation/is/research/projects.html

[123] Willinsky, Scholarly associations and the economic viability of open access publishing, in *Journal of Digital Information*, Volume 4, Issue 2, Article no. 177, 2003-04-09. http://jodi.tamu.edu/Articles/v04/i02/Willinsky

Impact of digital repositories on the economics of journal publishing

In addition to the Registry of Open Access Repositories (ROAR)[124], key background for this section is provided by the paper published in September 2005 by Van Westrienen and Lynch[125] (Source 6.8). The paper does not address economic issues, but is effectively a census with commentary, which demonstrates very well the proliferation of repositories in the countries studied. Also covered is the nature of the very diverse documentation which populates the repositories, including to a very considerable extent teaching material as well as research contributions. Except for the Netherlands (and possibly the USA, for which details were not available), the article also shows that the average number of items held in the repositories is still only 'a few hundred'.

Recently, the results of a study undertaken by Ware for ALPSP, which were published in March 2006 (Source 1.16, in Area 1)[126], have provided at least some initial data on the question of the possible linkage between the availability of self-archived articles in an OA repository and journal subscription cancellations by libraries. With 340 responses, admittedly from a self-selecting sample, the survey enjoyed a reasonably solid base, primarily from the USA and Western Europe. The key finding which informs the present study is that availability of articles in repositories was cited as either a 'very important' or an 'important' possible factor in journal cancellation by 54 per cent of respondents, even though ranking fourth after (i) decline of faculty need, (ii) reduced usage, and (iii) price. When respondents were invited to think forward five years, availability in a repository was still fourth-ranking factor, but the relevant percentage had risen to 81. Whilst this is not evidence of actual or even intended cancellation as a consequence of the growth of OA self-archiving repositories, it strongly suggests that such repositories are an important new factor in the decision process, and growing in significance.

Here, attention should also be drawn to the existence of several less formal communications relating to the possible effect of self-archiving repositories on journal usage and cancellation. In the fields of physics, astronomy, and mathematics, where repositories have the longest history, it has been reported that a number of learned societies, such as the Institute of Physics (IoP), the American Physical Society, and the London Mathematical Society (LMS), have experienced some decline in downloads from their journal sites when articles are also available in the long-established ArXiv repository[127]. For IoP Publishing, this trend was reconfirmed in a November 2005 presentation by John Haynes, Head of Business Development[128], while figures for the LMS are presented by Davis and Fromerth in a draft manuscript[129] which draws on data provided by the LMS. In this latter instance, Davis and Fromerth conclude that 'in general, articles

---

[124] ROAR – http://archives.eprints.org – currently lists 705 repositories

[125] Van Westrienen, G and Lynch, C, Academic institutional repositories: deployment status in 13 nations as of mid 2005, *D-Lib Magazine*, September 2005. http://www.dlib.org/dlib/september05/westrienen/09westrienen.html

[126] For a summary of the survey, see: Ware, M, Open archives and their impact on journal cancellations, in *Learned Publishing*, Vol. 19, no. 3, July 2006.

[127] See, for example the following related series of communications: (i) Swan (March 2005): http://www.keyperspectives.co.uk/openaccessarchive/Conference%20presentations/Presentation%20Berlin%203%202005.pdf (ii) Morris (August 2005) http://www.alpsp.org/2005pdfs/rcuk050805.pdf (iii) Lillywhite, then Harnad (September 2005): http://www.bio.net/hypermail/jrnlnote/2005-September/002749.html

[128] Haynes (November 2005): http://www.alpsp.org/events/2005/PPR/default.htm

[129] Davis P M, and Fromerth, M J. Does the arXiv lead to higher citations and reduced publisher downloads for mathematics articles? (updated May 2006) http://www.citebase.org/cgi-bin/fulltext?format=application/pdf&identifier=oai:arXiv.org:cs/0603056

deposited in the arXiv receive … fewer downloads at the publisher's website than non-deposited articles'.

The veracity of reports of reduced usage of publishers' journal sites (albeit in a narrow range of niche areas) is not doubted, and a causal link with availability in self-archiving repositories also appears to be accepted. There is, however, little or no evidence in relation to subscriptions, as well as disagreement over causality and the implications for peer review. Subscriptions are reported to have been declining over a period of 10+ years, but for a number of reasons (see discussion of Ware, above). Proving or disproving a link between availability in self-archiving repositories and cancellations will be difficult without long and rigorous research. In this connexion, the outcome of research recently announced by the Research Councils UK (RCUK)[130], with the co-operation of Macmillan, Blackwell and Elsevier, will be eagerly awaited, even though a report is not due until late 2008.

As regards authors' attitudes, the results of a recent survey by GfK NOP and Kindle Research for the Publishing Research Consortium[131] (Source 6.4) were reported in February 2006, and provide a sound body of data on author attitudes to the US National Institutes of Health (NIH) public access policy. A degree of author apathy and even scepticism is highlighted in the report. The findings, however, relate wholly to the USA and are no more than indicative of potential future difficulties in convincing UK authors to comply with comparable policies.

A slightly earlier study by Swan and Brown[132] (Source 6.6) covers similar ground, but is international in scope. The study had a solid base of nearly 1,300 respondents across a range of disciplines, of which 18 per cent were from the UK and a further 17 per cent from other EU countries. Overall, just under half of the respondents had had some experience of self-archiving, and over the previous 12 months the use of institutional repositories had doubled and the use of subject-based repositories had grown by 60 per cent. A very high proportion of authors (94 per cent) indicated that they would comply with a mandate to deposit in a repository, even if reluctantly in some cases.

Again, earlier research in this field has largely been superseded because of the speed of developments. Nevertheless, Ware, reporting in January 2004 (Source no. 6.9), is a useful round up of the position relating to repositories in general, including a review of initiatives and software. Ware also provides some useful data on publisher attitudes to repositories, based on 69 responses to an online questionnaire exercise carried out presumably in late 2003, as well as analysis of the contents of 45 institutional repositories.

A further publication from the Publishing Research Consortium, on the subject of the economic impact of repositories, had been expected in July/August 2006, but was not available for the present study, not having appeared by early September.

Analysis/data validation

For validation of data to be possible, it is essential to have established benchmarks, or at least to have datasets which have the same or similar terms of

[130] http://www.rcuk.ac.uk/access/2006statement.pdf
[131] Publishing Research Consortium/GFK/NOP, *NIH author postings: A study to assess understanding of, and compliance with, NIH public access policy*, February 2006. http://www.alpsp.org/news/NIH_authorpostings_report.pdf
[132] Swan, Alma, and Brown, Sheridan. *Open access self-archiving: an author study*. Key Perspectives Ltd, May 2005. http://cogprints.org/4385/

reference, with any non-overlapping areas clearly identified and their boundaries properly defined. This is not the case in several of the Areas discussed in this study, and the problem is particularly acute in Area 6. Such data as do exist are diverse in focus, and more often than not provide glimpses only, however fascinating, of limited areas of the topics under examination (as in, for example, close scrutiny of a small handful of journal titles, or examination of a narrow range of disciplines). Furthermore, in the debate over costs, as has emerged in Area 2 of this study, there are not even any solid comparators against which OA publication costs can be set.

This general paucity of sources means that most data are indicative rather than conclusive, and that establishing evidence-based causal relationships in key areas cannot currently be demonstrated. Similarly, extrapolation from restricted samples to wider communities is currently not possible. That there is little by way of solid data to analyse or validate is, therefore, a key finding of this study.

The only sub-area covered in this section of the study in which there may be said to be some agreement between sources, even though they do not wholly overlap in scope, is that of authors' attitudes. As already mentioned, this is not strictly within the terms of reference of Area 6, but authors' views will be an important factor affecting the cost and revenue bases of OA publishing, and thus will be one determinant of its success or otherwise. Two points stand out in particular. Firstly (and hardly surprisingly), there is strong support for the principle of OA in general (whether in OA journals or in self-archiving repositories), though that does not necessarily entail antagonism towards the existing subscription model. Secondly, there appears to be a willingness in general to comply with funders' mandates to deposit in repositories.

One other incipient development is worth noting. The focus of debates over costs seems to be gradually changing[133], although it is not at present possible to be entirely sure. There appears to be a degree of acceptance that many of the components of cost are common to both principal models and can therefore be cancelled out in the 'equation' – for example, 'first copy' costs at the beginning of the cycle, and server and software costs at the other end. There is currently little evidence available to assess the differences or similarities in other cost elements, e.g. overhead costs, launch costs, administrative costs in collecting author fees as opposed to subscriptions (and the cash flow implications), and the levels of surplus/profit generated.

Stemming from the above, there is greater concern as to how the burden of author funding will be borne, given that a straightforward institutionally-based solution would potentially be inequitable, concentrating the cost on a relatively small number of research-intensive institutions. This is exactly the issue highlighted above in discussing Dominguez (Source 6.1).

Gaps in the data

The specification for this study articulates the basic tasks as follows:

- To find out what data and sources can be relied on;
- To identify gaps;

[133] For example, c.f. Jan Velterop's comment: "Let's call a spade a spade, then. It's not about OA, because that's clearly accepted as beneficial. It's about who and how OA is paid for. A legitimate discussion that should not be blurred by arguments that seem to be about OA *per se*." https://arl.org/Lists/SPARC-OAForum/Message/2438.html

- To identify widely-believed but inaccurate data.

Taking the first bullet point, the answer in summary is that there are some useful studies and solid data, but that, as already mentioned above there is a patchwork of niche studies which are too narrow in scope to provide a basis for general conclusions or extrapolation.

As regards the third point, in a situation where little has been incontrovertibly proved, much that is believed could be wrong, but the absence of evidence makes rebuttal difficult. Open minds are, therefore, essential in this situation pending the completion of well-founded research designed to permit a fuller understanding of the OA phenomenon.

The other key objective is to identify gaps. In Area 6, however, there are considerable voids with a few oases of data, so making recommendations for specific projects is difficult. In general terms, therefore, it seems sensible to say that the focus needs to be on issues that are particularly contentious or otherwise difficult. Key areas which suggest themselves (and which might possibly be combined) are:

- More work is needed on the relationship (if any) between subscription cancellations and repositories. The work will need to be carefully designed, with sufficient, representative and balanced samples, and the collaboration of all stakeholders, including especially research institutions and publishers. Any such study will also need to be maintained over a fairly extended period, with regular reports, since it seems likely that the position could change with time if the contents of self-archiving repositories become progressively more comprehensive.
- Similarly, more carefully conceived work on the impact of both OA journals and self-archiving on the quality of research communications, especially on the peer review system.
- A wide-ranging study on the funding and money flow implications of new publishing paradigms, particularly if the author-pays model becomes well established.

Possibly, the project planned by RCUK[134], already mentioned above, will address both issues, although the precise terms of reference are not known at this stage.

Returning again to the specification, this study was asked to consider the following:

What is known or believed?
- What can be proved?
- What cannot be proved?

What is uncertain?
- What is capable of being proved?
- What cannot be known for certain?

In relation to what is currently known or believed, the discussion in the preceding sections and the detailed evaluation sheets in Appendix 1 have sought to highlight the degree to which propositions are underpinned by evidence or not. The other three questions can perhaps best be taken together. In principle, it should be possible to construct research which will test hypotheses concerning any market phenomena, provided that the right questions can be asked and relevant factors isolated. The difficulties lie in obtaining an adequate response –

[134] See: http://www.rcuk.ac.uk/access/2006statement.pdf

adequate not just in numbers, but in quality of response and in breadth of coverage. This requires a serious objective effort on the part of respondents. These questions can only be resolved with a concerted effort on the part of all involved in the marketplace. This is not impossible, but does demand major effort, commitment, and robust financial support.

Sources evaluated in detail in Appendix 1

Source 1: Dominguez, Magaly. Economics of open access publishing, in *Serials*, 19(1), March 2006. http://serials.uksg.org/(pguu4jmkt5y0kc45gjzblr3k)/app/home/contribution.asp?referrer=parent&backto=issue,14,20;journal,1,34;homemainpublications,2,2;

Source 2: Hedlund T, Gustafsson T, Björk B. The open access scientific journal: an empirical study, July 2004. http://oacs.shh.fi/publications/199-210.pdf

Source 3: Kaufmann-Wills Group LLC on behalf of ALPSP, AAAS, HighWire Press and AAMC. *The facts about open access: a study of the financial and non-financial effects of alternative business models for scholarly journals* [Including post-publication addendum with additional, corrected data and analysis on peer-review] (October 2005). http://www.alpsp.org/publications/FAOAcompleteREV.pdf

Source 4: Publishing Research Consortium/GFK/NOP, *NIH author postings: A study to assess understanding of, and compliance with, NIH public access policy*, February 2006. http://www.alpsp.org/news/NIH_authorpostings_report.pdf

Source 5: Swan, Alma and Brown, Sheridan. Authors and open access publishing, *Learned Publishing*, Vol. 17, No. 3 (July 2004), pp 219-24. http://puck.ingentaconnect.com/vl=19271709/cl=11/nw=1/fm=docpdf/rpsv/cw/alpsp/09531513/v17n3/s7/p219

Source 6: Swan, Alma, and Brown, Sheridan. *Open access self-archiving: an author study*. Key Perspectives Ltd, May 2005. http://cogprints.org/4385/

Source 7: SQW Ltd for the Wellcome Trust, *Costs and business models in scientific research publishing*, April 2004. http://www.wellcome.ac.uk/assets/wtd003184.pdf

Source 8: Van Westrienen, G and Lynch, C, Academic institutional repositories: deployment status in 13 nations as of mid 2005, *D-Lib Magazine*, September 2005. http://www.dlib.org/dlib/september05/westrienen/09westrienen.html

Source 9: Ware, M. *Pathfinder Research On Web-Based Repositories: Final Report*, January 2004. http://www.palsgroup.org.uk/palsweb/palsweb.nsf/0/8c43ce800a9c67cd80256e370051e88a/$FILE/PALS%20report%20on%20Institutional%20Repositories.pdf

Source 10: Willinsky, Scholarly associations and the economic viability of open access publishing, in *Journal of Digital Information*, Volume 4, Issue 2, Article no. 177, 2003-04-09.
http://jodi.tamu.edu/Articles/v04/i02/Willinsky

Bibliography

ALPSP & CAPP (2002). Learned Journal Publishing in the UK. Briefing Prepared for UK Government.

Antelman K, (2005). Do open access articles have greater research impact? College and research Libraries, 65 (1), 372-382.

Baldwin, C (2004). What do Societies do with their Surpluses? ALPSP/Blackwell Survey.

Bailey, J (1989). New Journal Decision Making. College & Research Libraries, May, 1989.

Bauer K and Bakkalbasi N, (2005). An examination of citation counts in a new scholarly communication environment. D-Lib Magazine, 11 (9). http://www.dlib.org/dlib/september05/bauer/09bauer.html

Bergstrom, T & McAfee, P. Journal Cost Effectiveness Summary Statistics. www.hss.caltech.edu/~mcafee/Journal/Summary.pdf

Bornmann L and Daniel H (2005), Selection of research fellowship recipients by committee peer review. Scientometrics, 63 (2), 297-320.

Brody T and Harnad S, (2006). Earlier Web usage statistics as predictors of later citation impact, http://eprints.ecs.soton.ac.uk/10713/01/timcorr.htm

Butler, D, (20 June 2006). Open-access journal hits rocky times, in Nature. http://www.nature.com/news/2006/060619/full/441914a.html

Competition Commission (July 2001). Reed Elsevier plc and Harcourt General, Inc, A Report on the Proposed Merger. http://www.competitioncommission.org.uk/rep_pub/reports/2001/fulltext/457c1.pdf

Conyers, A. and Dalton, P. (2005). Evidence Base. NESLi2 analysis of usage statistics: summary report.

Cox J and Cox L (2003 and 2005). Scholarly Publishing Practice – The ALPSP report on academic journal publishers' policies and practices in online publishing (First and Second Surveys).

Craig I, (August 2006). The effect of self-archiving and Open Access on citations. Blackwell Scientific Internal Report.

Creaser C, (2006). Evaluation of the open access journal experiment: stage 2 report, in Oxford Journals, Assessing the impact of Open Access. Oxford University Press.

Creaser, C, Maynard, S, White, S (November 2005). LISU Annual Library Statistics 2005.

Cronin B and Meho L (2006). Using the h-index to rank influential information scientists, Journal of the American Society for Information science and Technology, 57 (9), 1275-1278.

CSA, Ulrich's Periodicals Directory.

Davis, P M, and Fromerth, M J, (updated May 2006). Does the arXiv lead to higher citations and reduced publisher downloads for mathematics articles? http://www.citebase.org/cgi-bin/fulltext?format=application/pdf&identifier=oai:arXiv.org:cs/0603056

Davis, P (Revised December 2005). Journal Cost Effectiveness Tables and Graphs. http://people.cornell.edu/pages/pmd8/

Davis, P and the Cornell University Library Task Force on Open Access Publishing, (Dec 22, 2004). Calculating the Cost per Article in the Current Subscription Model. http://people.cornell.edu/pages/pmd8/

Debackere K and Glanzel W, (2004). Using a bibliometric approach to support research policy-making. Scientometrics, 59 (2), 253-276.

Department of Trade and Industry: Publishing Sector: www.dti.gov.uk/sectors/publishing/index.html

Dewatripont, M, Ginsburgh, V, et al. for the European Commission (January 2006). Study on the economic and technical evolution of the scientific publication markets in Europe, D-G Research. http://ec.europa.eu/research/science-society/pdf/scientific-publication-study_en.pdf

Dominguez, Magaly (March 2006). Economics of open access publishing. Serials, 19(1). http://serials.uksg.org/(pguu4jmkt5y0kc45gjzblr3k)/app/home/contribution.asp?referrer=parent&backto=issue,14,20;journal,1,34;homemainpublications,2,2

Donovan, B (1998). The Truth About Peer Review, Learned Publishing, 11(3), 170-184.

Dryburgh, A (2002). The costs of learned journal and book publishing: a benchmarking study for ALPSP.

Education for Change, SIRU, and Research Partnership, (2002). Researchers use of libraries and other information sources.

eJust (e-Journal user study), (2002). Stanford University.

Electronic Publishing Services Ltd., (January 2005). Market Monitor: Education and Training: market trends and industry performance.

Electronic Publishing Services Ltd., (May 2005). Market Monitor: Legal, Tax and Regulatory (LTR) Information: market trends and industry performance.

Eysenbach G, (2006). Citation advantage of open access articles. PLoS Biology. http://biology.plosjournals.org/perlserv/?request=get-document&doi=10.1371/journal.pbio.0040157

Hajjem C, Harnad S and Gringras Y, (2005). Ten-Year cross-disciplinary comparison on the growth of Open Access and how it increases research citation impact. IEEE Data engineering Bulletin, 28 (4), 39-47.

Harnad S et al (2004). Comparing the impact of open access versus non open access articles in the same journals. D-Lib magazine, 10 (6). http://www.dlib.org/dlib/june04/harnad/06harnad.html

Haynes (November 2005). http://www.alpsp.org/events/2005/PPR/default.htm

Hedlund T, Gustafsson T, Björk B (July 2004). The open access scientific journal: an empirical study. http://oacs.shh.fi/publications/199-210.pdf

Herbertz, H, (1995). Does it pay to cooperate? A bibliometric case study in molecular biology, Scientometrics, 33 (1), 117-122

Hitchcock S, (2006). The effect of Open access downloads on citation impact: a bibliography of studies, http://opcit.eprints.org/oacitation-bibio.html

House of Commons Science and Technology Committee. Scientific Publications: Free for all? Volume II: Oral and Written Evidence. (HC 399-II). Page 27: Response to Question 188.
http://www.publications.parliament.uk/pa/cm200304/cmselect/cmsctech/399/40 30808.htm

Kaufmann-Wills Group LLC on behalf of ALPSP, AAAS, HighWire Press and AAMC (October 2005). The facts about open access: a study of the financial and non-financial effects of alternative business models for scholarly journals [Including post-publication addendum with additional, corrected data and analysis on peer-review]. http://www.alpsp.org/publications/FAOAcompleteREV.pdf

Keenan S. and Johnston C., (2000). Concise Dictionary of Library and Information Science.

Key Perspectives, (1999). What authors want: the ALPSP research study on the motivations and concerns of contributors to learned journals.

Key Perspectives, (2002). Authors and electronic publishing: the ALPSP research study on authors and readers views of electronic research communications.

Key Perspectives, (2004). Authors and open access publishing. Learned Publishing, 17(3), 219-224.
http://puck.ingentaconnect.com/vl=19271709/cl=11/nw=1/fm=docpdf/rpsv/cw/alpsp/09531513/v17n3/s7/p219

Key Perspectives Ltd. for JISC/OSI, (2004). JISC/OSI Journal authors survey.

Key Perspectives Ltd., (May 2005). Open access self-archiving: an author study. http://cogprints.org/4385/

King, D & Tenopir, C, (June 2004). An evidence-based assessment of the 'author pays' model, Nature Web Focus.
http://www.nature.com/nature/focus/accessdebate/26.html

Kurtz M.J et al (2005). The effect of use and access on citations, Information Processing and Management, 41, 1395-1402.

Lawrence S. (2001). Free online availability substantially increases a paper's impact, Nature, 411, 521.

Loke, Y.K. and Derry, S. (2003). Does anybody read 'evidence based' articles?

Lynch, C, (February 2003). Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age. ARL bimonthly report, 226, 1-7. http://www.arl.org/newsltr/226/ir.html

Mabe M.A. and Amin M. (2002). Dr Jekyll and Dr Hyde: author-reader asymmetries in scholarly publishing. Aslib Proceedings: new information perspectives, Volume 54, Number 3, August 2002, pp. 149-157(9). Emerald Group Publishing Limited.
http://www.ingentaconnect.com/content/mcb;jsessionid=24kqc7s7c19lk.alice
http://www.ingentaconnect.com/content/mcb/276;jsessionid=24kqc7s7c19lk.alice

Mabe M & Amin M, (2001). Growth Dynamics of Scholarly and Scientific Journals. Scientometrics, Vol. 51 (1).

Mabe, M (2003). The Growth and Number of Journals. Serials, Vol. 16, No. 2.

Mellman I, (July 2006). How Journals Can Realistically Boost Access. Nature Web Focus.

Moed H.F, (2005). Citation analysis in research evaluation, Springer.

Morris, S, (2006). When is a journal not a journal? A closer look at the DOAJ. Learned Publishing, Vol 19 (1), January 2006.

Morris, S, (2005). The True Costs of Scholarly Journal Publishing. Learned Publishing, Vol 18 (2), April 2005.
http://alpsp.publisher.ingentaconnect.com/content/alpsp/lp/2005/00000018/00000002/art00006;jsessionid=2p98rnbw2m1ne.victoria#avail

Nicholas, D. et al (2005). The big deal – 10 years on.

Nicholas, D et al (2005). Scholarly journal usage: the results of deep log analysis.

Norris M and Oppenheim C (2003). Citation counts and the Research Assessment Exercise V: Archaeology and the 2001 RAE, Journal of Documentation, 59 (6), 709 – 730, and references cited therein.

Oxford Journals (2006). Assessing the impact of Open Access, Oxford University Press.

Oxford Journals, Oxford University Press (June 2006). Assessing the impact of open access – preliminary findings from Oxford Journals.
http://www.oxfordjournals.org/news/oa_report.pdf

Page, G, Campbell, R, Meadows, J (1997). Journal Publishing. Cambridge University Press, pp 407 ISBN 0 521 44137 4.

Publishing Research Consortium/GFK/NOP, NIH author postings, (February 2006). A study to assess understanding of, and compliance with, NIH public access policy. http://www.alpsp.org/news/NIH_authorpostings_report.pdf

Rightscom Ltd (2005). JISC Disciplinary differences report.

Rowlands, I. and Nicholas, D (2005). New Journal Publishing Models.

Rowlands, I., Nicholas, D., and Huntingdon, P, (2004). Scholarly communication in the digital environment: what do authors want?

Rowlands, I and Olivieri, R, (2006). Journals and scientific productivity.

Rowland, F (2002). Report for the JISC Scholarly Communications Group. The Peer Review Process.

Schonfeld, King, Okersen and Fenton (2004) on behalf of the Council on Library and Information Resources. The Non-Subscription Side of Periodicals: Changes in Library Operations and Costs between Print and Electronic Formats.

Shadbolt N, Brody T, Carr L and Harnad S, The open research web, in N. Jacobs (Editor), Open Access: key strategic, technical and economic aspects, Chandos Publishing (2006), 195-208.

SQW Ltd for the Wellcome Trust (April 2004). Costs and business models in scientific research publishing. http://www.wellcome.ac.uk/assets/wtd003184.pdf

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

Tenopir, C & King, D (December, 1998). Designing Electronic Journals with Thirty Years of Lessons from Print, The Journal of Electronic Publishing, Volume 4, Issue 2.

Tenopir, C *et al.* (2003), Patterns of journal use by scientists through three evolutionary phases.

Tenopir, C & King, D (1998). Economic Cost Models of Scientific Scholarly Journals, University of Tennessee School of Information Sciences Center for Information Studies Paper presented to the ICSU Workshop, Keble College, Oxford, UK, 31 March to 2 April 1998.

The Office of Science and Innovation (DTI) (November 2005). Science, engineering and technology indicators.
http://www.dti.gov.uk/science/science-funding/set-stats/gross-domestic-exp/index.html

Thomson ISI, (2004). The impact of open access journals – a citation study from Thomson ISI. http://scientific.thomson.com/ts/media/presentrep/acropdf/impact-oa-journals.pdf

Thomson Scientific Journal Selection Process (last updated January 2004): http://scientific.thomson.com/free/essays/selectionofmaterial/journalselection

Thomson Scientific (ISI). Journal Citation Reports and Indexes.

UNESCO (2005). UNESCO Science Report: http://www.uis.unesco.org/

University of Wisconsin, Madison Campus Libraries: 'Journal Value Project (JVP)'. http://www.wendt.wisc.edu/projects/jvp/welcome.do

Van Orsdel, LC & Born, K, (2005). Library Journal, Periodical Price Survey, April 15.

Van Raan, A. F. J (1998). The influence of international collaboration on the impact of research results: Some simple mathematical considerations concerning the role of self-citations, Scientometrics, 42 (3), 423-428.

Van Westrienen (September 2005). G and Lynch, C, Academic institutional repositories: deployment status in 13 nations as of mid 2005, D-Lib Magazine. http://www.dlib.org/dlib/september05/westrienen/09westrienen.html

Vogel E.E. (1997). Impact Factor and international collaboration in Chilean physics, 1987-1994, Scientometrics, 38 (2), 253-263.

Waltham, M, on behalf of JISC (June 2005). Learned Society Open Access Business Models

Ware, M, (January 2004). Pathfinder Research On Web-Based Repositories: Final Report.
http://www.palsgroup.org.uk/palsweb/palsweb.nsf/0/8c43ce800a9c67cd80256e370051e88a/$FILE/PALS%20report%20on%20Institutional%20Repositories.pdf

Ware, M, on behalf of ALPSP (March 2006). Survey of librarians on factors in journal cancellation.

Ware, M (July 2006). Open archives and their impact on journal cancellations, in Learned Publishing, Vol. 19, no. 3.

White, S, & Creaser, C, of LISU, (2004). Scholarly Journal Pricing: Selected Trends and Comparisons. LISU, University of Loughborough.

Willinsky, J, (2003). The Nine Flavours of Open Access Scholarly Publishing. J Postgrad Med, 49:263-267: http://www.jpgmonline.com/article.asp?issn=0022-3859;year=2003;volume=49;issue=3;spage=263;epage=267;aulast=Willinsky

Willinsky, J, (9 April 2003). Scholarly associations and the economic viability of open access publishing, in Journal of Digital Information, Volume 4, Issue 2, Article no. 177. http://jodi.tamu.edu/Articles/v04/i02/Willinsky

Index of tables