

UK scholarly journals: 2006 baseline report

An evidence-based analysis of data
concerning scholarly journal publishing

Executive Summary

The full version of the report, along with
an electronic copy of this summary, can be found at:
<http://www.rin.ac.uk/data-scholarly-journals>

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)



www.rin.ac.uk



www.rcuk.ac.uk



www.dti.gov.uk

Executive summary

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 six key areas of:

1. *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
2. *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
3. *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
4. *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
5. *disciplinary differences*, and whether there are significant differences between researchers as readers or authors in different disciplines
6. *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.

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, the available evidence is largely 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 this 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>

