A commentary on 'The economic implications of alternative publishing models'

Steven Hall *

* Steven Hall Consulting Limited, Dorchester on Thames, UK

Online publication date: 24 March 2010
RESPONSE

A commentary on ‘The economic implications of alternative publishing models’

Steven Hall*

Steven Hall Consulting Limited, Dorchester on Thames, UK

Introduction

In ‘The economic implications of alternative publishing models’, Houghton and Oppenheim summarise a much longer and more detailed report (Houghton et al., 2009) published by the Joint Information Systems Committee (JISC) in January 2009. This original report piled assumption on assumption, estimate on estimate, to arrive at a series of conclusions about the potential economic benefits of open access publishing which have been widely quoted by proponents of open access, but which are deeply flawed. This commentary reviews these assumptions and estimates to show that the conclusions drawn from them about the savings and benefits to be gained from open access publishing over traditional publishing models are wrong. As the devil is in the detail, the commentary refers frequently to the original report.

Alternative publishing models

Houghton and Oppenheim looked at three scholarly publishing models: the subscription publishing model, in which the publisher charges a fee for a subscription to a journal or the purchase of a book; the open access publishing model, in which access to the journal or book is free of charge with publication costs being paid by the author or the author’s institution or funding body; and the open access self-archiving model, under which the author deposits a manuscript in a freely accessible online repository. They acknowledge that the last is not in itself a formal publishing model, but they seek to turn it into a formal model by either running self-archiving in parallel with subscription publishing, or overlaying on it some form of post-archiving peer review, quality control and branding.

A key difficulty in comparing these models lies in the fact that while the subscription model is mature and costs can be known (though the report’s authors do not have access to real cost data), the open access model is immature and not yet proven to work in any sustainable or scalable way. The self-archiving model is impossible to run in parallel with the subscription model, and there is no working example on any scale whatsoever of their overlay model. Yet this does not stop them from comparing the three (or four), as though all costs for each are known and reliable. Indeed, they claim that their figures for the open access model include full commercial margins, while ignoring the fact that they do not know whether commercial publishers are making

*Email: slcshall@btinternet.com
any margin at all from such open access models, and that the largest open access publishing operation which does publish its financial results, the Public Library of Science, is yet to break even on an annual basis and may never do so on a cumulative basis, in spite of receiving close to US$12m in grant funding between 2004 and 2008. For this reason the use of £1500 as the publication fee under the open access model needs much closer scrutiny than it is given in the report. The authors cannot simply assume that this includes a ‘full commercial margin’.

**Approach and methodology**

This commentary will not address the model of the scholarly communications life cycle used in the report. It must, however, take issue with the sources used by the authors for their data. There was no engagement whatsoever with the subscription publishing industry during the development of the report. JISC (2009) has claimed that this was because the report was intended to be independent. This appears to mean that it should be independent of the subscription publishing industry, but not of the open access lobby, which was strongly represented in the report’s authors and expert review group. The original report has not undergone standard peer review.

**Lack of real-world data**

The lack of involvement of subscription publishers in the report required the authors to rely on a number of academic studies. One such study heavily referenced by Houghton and Oppenheim is Halliday and Oppenheim (1999). 1999 is a very long time ago in digital publishing, only two or three years into the online publishing of journals and before any significant migration from print to online. It is well before the advent of the Big Deal (beyond early experiments such as the UK Pilot Site Licence Initiative), Google Scholar, Scopus and the ebook, to mention just four key developments since then. The research is out of date. It is also unrealistic for the authors to demand that publishers respond to the report by inputting their own cost data into the model and publishing the results; as they well know, such data are extremely commercially sensitive in a highly competitive industry. The authors claim to have supplemented these academic studies ‘where necessary’, by consulting experts in the field. Without doubt, the experts in the field on publishing costs are scholarly publishers, but they were not consulted.

**Selective use of data**

The report consistently overstates the case for open access publishing and understates the case for subscription publishing, in its handling of its estimates of the costs and benefits of each, in its discussion of the citation impact of open access, and even in its coverage of access to research literature in the developing world. In relation to this last point, it states that ‘a number of authors have noted the particular benefit of open access for developing countries, where access to the subscription-based literature has often been limited’. This statement is supported with three anecdotes, but no other evidence. The report fails to mention the Research4Life project, under whose umbrella more than 130 science publishers, the WHO, FAO and UNEP, Cornell and Yale Universities and their technology partner, Microsoft, are collaborating to make
more than 7500 peer-reviewed scientific journals, plus many books, indexes and databases, available free of charge or at very low cost to researchers at 4500 institutions in the developing world. These are the HINARI, AGORA and OARE projects (the first running since 2002), from which researchers in the developing world downloaded more than 6.5 million articles in 2008. While no-one would claim that there is no knowledge gap between the industrialised world and the developing world, for the report to publish three anecdotes in support of its own case and to make no mention of the widely publicised efforts that publishers and others are making to bridge that gap reflects its tendentious nature.

Costs and benefits
The estimated savings and benefits of a move from subscription publishing to one or other form of open access publishing identified by Houghton and Oppenheim can be assigned to three main categories: direct savings in publishing costs; indirect savings in library costs and the costs of research performance; and gains in the return to investment in R&D. One of the difficulties in responding to the report is its intermingling of the benefits of a move from today’s combination of print and online publishing to a future which is online-only, with the estimated benefits from a move from subscription publishing to open access publishing. The authors claim that it is difficult to separate the two:

One of the keys to comparing the costs of alternative publishing models is to disentangle the cost impacts of format (i.e. print versus electronic) and model (i.e. toll versus open access). This is very difficult to do. (Houghton et al., 2009, p. 165)

Is it really so difficult? All publishers are modelling the cost differences as their businesses move from today’s mix to a largely online future. The report’s methodology is wrong. It would have been perfectly straightforward and more transparent to have done this work in two separate stages, as the CEPA (2008) study did, first looking at the impact of a move to online only, and then looking at what additional benefits, if any, might accrue from a move from subscription publishing to open access publishing. Instead, the report mixes the two up. In places it manages to separate them, but in far too many other instances it conflates them, ascribing both direct and indirect savings to a move to open access publishing when rightly they result from a move to online-only publishing. Publishers would generally be happy to move from the current mixed model to an online-only model. What they cannot do, though they have lobbied hard, is deal with the VAT issue in the European Union. The single largest barrier to a faster move to online-only journal publishing is the much higher level of VAT on online journals.

Publishing cost savings
There is not the space here to address all the flawed assumptions on the relative costs of subscription and open access publishing in the report; they are myriad. We will look at just four examples. First, there is the comparison of the cost of processing payments from institutions for their journal subscriptions with that of processing the payment of publication fees by authors. The report compares the cost of managing payments for tens of thousands of subscriptions to individual journals with the cost of
managing payments from tens of thousands of article authors and produces a saving of 80% on these costs. It identifies publishing costs of £100 per article under the subscription model, and of £20 per article under the open access model. (It claims this is based on the CEPA study, though CEPA in fact calculated a cost of £30 per article for processing author-side payments. Why the report reduced this by a third is not explained.) What the report misses is the fact that publishers are not invoicing libraries for thousands of journals individually, but, rather, are issuing a single invoice in respect of all a library’s subscriptions or its Big Deal. In many cases, they are doing this through a subscription agent, further simplifying the process. The cost of processing nearly 100,000 author payments, in the case of UK authors only, will far exceed the cost of invoicing a couple of hundred UK higher education institutions. Furthermore, the report takes no account of the cost to authors of managing such payments, nor of the cost to their institutions or funders of administering author payments.

Second, there is an equally unsupported claim for a saving of 67% on marketing costs:

Drawing on a range of sources, we estimate marketing costs at £120 per article for the subscription model and a conservative £40 per article for the OA publishing model (i.e. marketing to authors). (Houghton et al., 2009, p. 155)

The authors fail to say which sources they drew on, and in what way their open access cost of £40 is ‘conservative’, but they clearly were not in touch with the reality of journals publishing. The largest single marketing expense for any journals publisher is attendance at academic conferences. The main purpose of that attendance is to attract authors to the publisher’s journals. Another substantial part of a journal publisher’s marketing expenditure is promoting usage, through indexing, search and discovery tools, promotion of articles to the user community, and so on. A relatively small part of a journal publisher’s marketing budget is spent on marketing to libraries; far more on marketing to authors and users.

This saving is based on yet another false assumption, that less marketing would be necessary with open access. There are far more authors than institutional customers and there would be even more intense competition for authors, which would require greater promotion to them and even greater investment in author services.

Similar claims are made for savings of 50% on online hosting:

Following CEPA (2008) we estimate online hosting costs per article at £200 for the subscription model, and £100 for the OA publishing model – with less use of proprietary access systems and no need for access control and authentication in the latter. (Houghton et al., 2009, p. 155)

This demonstrates again a fundamental lack of understanding of how online scholarly publishing works today and how it would be likely to work under an open access model in which there were competition between providers. Publishers will want to differentiate their services from those of their competitors, through value-added services to users, authors, reviewers, editors, and so on. They will not want simply to deposit articles on a platform like PubMed Central. The reference to ‘proprietary access systems’ here also shows another misunderstanding of how publishers deliver their online journals today. They do not all use ‘proprietary’ systems. A good number use third-party platforms like HighWire; they choose their platform on the basis of functionality, cost and other similar measures. They do not choose HighWire
because it is inexpensive; as a highly functional platform, it is not. There is no saving of 50% to be had. Authentication is a tiny part of the cost of running a robust online platform. If it went away, savings would be negligible. Furthermore, there would have to be additional investment in systems; for example, to understand usage better, which would need to be captured and reported on at author level.

As a final example, there is an estimated saving of 25% on management and investment margins, achieved through ‘more predictable and stable’ revenue streams and payment by authors rather than institutions. The idea that revenue would be more predictable under an open access model is highly questionable. One of the big attractions for those who invest in the scholarly publishing industry today is the predictability and stability of subscription income. Even more questionable is the idea that investment will be reduced as ‘author fees materialise immediately’. The report’s authors fail to explain in what way the payment of author fees shortly in advance of publication could compare favourably with the advance payment of annual subscriptions, up to six months before the start of the subscription cycle. The cash flow implications of the open access model are negative, not positive. Yet again, the estimated saving here is a fantasy based on a profound misunderstanding of how scholarly publishing works.

There are similar unsupported claims for cost savings or benefits in the areas of online user management, rights management, customer service and help desk, operating margins and even advertising revenue. Rather than look at each of these, let us simply review the total savings estimated by the authors:

Hence, on average estimated costs, a shift from all toll access e-only to OA e-only publishing for all journal articles produced in UK higher education during 2007 would have directly saved around £80 million, and for authored and edited books around £94 million. A shift from all toll access e-only to OA self-archiving e-only with overlay services for all journal articles produced in UK higher education during 2007 would have saved around £116 million (an additional £36 million), and for authored and edited books around £102 million (an additional £8 million). (Houghton et al., 2009, p. 184)

So, Houghton and Oppenheim have estimated direct savings of £80m on journals and £94m on books for UK higher education from a move to open access publishing, and of £116m on journals and £102m on books from a move to self-archiving with overlay services.

Where are these savings to be realised? They must be set against UK university library spending in 2007, according to the Society of College, National and University Libraries (SCONUL) figures, of £113m on serials and £56m on books (Table 1). So, under the open access model the report claims that the UK would save 71% of its current expenditure on journals without any subscription cancellations, and would save £38m more than it is actually spending on books. Under the self-archiving model,

| Table 1. UK Library expenditure on books and journals and estimated savings |
|------------------|-----|-----|
|                  | Serials | Books |
| SCONUL expenditure 2007 | £113m | £56m |
| Houghton and Oppenheim direct savings from open access model | £80m | £94m |
| Houghton and Oppenheim direct savings from self-archiving model | £116m | £102m |
the UK would save £3m more than it is currently spending on journals, again without any cancellations, and £46m more than it is spending on books.

This reflects, of course, the problem of developing a theoretical model with little or no relation to the real world. It is also the problem of trying to build a model up from the estimated cost of a single article or book. You make errors in your assumptions, gross the whole thing up and your numbers simply do not match anything that anyone operating in that real world would recognise. It also reflects the problem of trying to look at the cost of UK research output against UK spending on global research output. You cannot match up the numbers. But it must be clear from the most cursory examination that these are theoretical and not real savings. You cannot save more than you are currently spending.

Library and research performance savings

The report estimates potential savings for UK higher education libraries of £34m from a move to online-only publishing, and an additional £11m from the move from subscription publishing to open access publishing. The response to the report from the Publishers Association, the Association of Learned and Professional Society Publishers (ALPSP) and the International Association of Scientific, Technical and Medical Publishers (STM), pointed out that to realise the savings of £11m, more than 200 librarian jobs would have to be lost. The JISC (2009) response to the publishers’ comments countered that ‘savings realized would release resources to more research and research support activities, and would not be clawed back in funding cuts’. You cannot have your cake and eat it. You either realise savings or you do not.

The report is also highly equivocal about the role that libraries play in enabling easy access to the right content for their users:

OA e-only journal handling expenditure could be considered discretionary, as user communities could discover and access the material independent of their research libraries. However, it is included to provide a basis for cost comparisons between publishing models. (Houghton et al., 2009, footnote to p. 170)

On the one hand, the report proposes massive savings in research performance through the easier discoverability of relevant content in an open access world; on the other hand, it suggests a kind of free-for-all in which users are left to their own devices to find the content they need. Once again, you cannot have it both ways. Libraries play an essential role in providing access to the content that their specific user communities need and will do so regardless of the business model under which it is published; the report fails to understand this.

Research performance savings and increased returns to investment in R&D

In looking at publishing and library costs, we have been dealing largely with figures that could be known and extrapolated from, even if the report failed to get many of them right. But as we move from looking at publishing costs to looking at the estimated savings in research performance, and then further to the estimated increase in returns to investment in R&D, we move from the non-fiction shelves towards the (science) fiction.
The estimates for research performance savings and increased returns to R&D are addressed here together as they are both based on the same assumptions of the benefits of improved access to research information. The report claims that annual savings of £108m could be realised in the UK on research performance and research funders’ costs, through speedier access to scientific information; and that a gain of £329m could be achieved in the annual return to investment in R&D. Together, these savings and returns dwarf the savings on publishing and library costs.

The research performance and funder savings are said to be realisable through savings in the time of funding bodies, reviewers and researchers, based on estimates of time saved in various tasks of between 5% and 50%. The estimated improvement in returns to R&D investment are based on a 20% return to R&D and a 5% improvement on that return, again through easier access to scientific research (including in the developing world, where we have already seen that the report is misleading on current access provisions).

**Access to research literature**

This is all largely predicated first on researchers and others having access to less than 50% of the research output they need:

Hence, as a simple proxy, perhaps 50% of possible journal titles are not readily accessible to higher education researchers in the UK. (Houghton et al., 2009, p. 201)

The report assumes that 50% of journals equal 50% of articles. This is simply wrong. Only 33% of journals account for 80% of all articles published, and 50% of journals account for 90% of all articles published (Figure 1).

![Figure 1. Journal article distribution](source: Scopus)
The open access citation effect

The second driver of these savings is the supposed citation advantage of open access publishing:

… as [a] starting point one might take 25% as a conservative estimate of the potential citation advantage of OA publishing models. (Houghton et al., 2009, p. 202)

This ‘conservative’ estimate is based yet again on a partial reading of the literature. Four important studies (Craig et al., 2007; Kurtz and Henneken, 2007; Moed, 2007; Davis et al., 2008), all published well before the report and therefore available to its authors, are not even listed in its bibliography, never mind taken into consideration in its findings. They all conclude that there is no citation effect for open access publishing. To quote from just one of them:

There are a number of excellent arguments in favor of changing the scientific publication system to an open access model. The open access citation advantage is not one of them. (Kurtz and Henneken, 2007)

In 2009, two further studies of the so-called citation impact were published showing no citation effect of open access publishing in the disciplines of Economics (Frandsen, 2009) and Ophthalmology (Lansingh and Carter, 2009). Once again, we have to conclude that the authors’ interpretation of the published research is flawed. At best, the jury is still out on the effect of open access on citations, though the most recent research is showing no effect at all, across a wide range of disciplines.

The fact is, researchers today have immediate access to the vast majority of the scientific articles that they need for their research. This is thanks to two things, the first being years of fine-tuning of collections by librarians to ensure that their users have access to the core journals for their disciplines. The second is the impact of the Big Deal, under which the number of journals to which the average academic researcher in the UK has immediate access via his or her university library has more than doubled over the last 10 years.

In the companion report to a study published by the Publishing Research Consortium (Ware, 2009), 94% of academic respondents to a survey on access to journal literature reported very easy or fairly easy access. In the same survey, academic users placed access to journal literature 13th out of 16 in a list of barriers to success at their institution. These are not statistics which support a case for difficulty of access to the literature. Of course, the Big Deal doesn’t provide every researcher at every institution with access to every journal, but an institution with no medical school does not need the vast majority of medical journals and the LSE does not need journals in high-energy physics.

The economic impact of the Big Deal

If the marginal additional access to the literature provided by open access produced substantial savings in research performance plus a 5% improvement in research efficiency, with a total annual value to UK higher education of £437m, then what savings and improvement in research efficiency has the Big Deal provided? We know that it has more than doubled the number of journals to which UK researchers
have access and we know that it is well used. Using the report’s assumptions, one might conclude that the Big Deal would be worth a gain of at least 10% in research efficiency and probably more than £1 billion to the UK economy. For an investment of probably under £10m (the JISC should have the exact figures from its NESLi negotiations) we have a 100-fold gain, not that publishers make such extravagant claims. In the real world, the Big Deal exists and presumably its impact on research performance and efficiency can be measured. If Houghton and Oppenheim can show the value of the Big Deal in these terms, then perhaps we can extrapolate from that value an additional value for the marginal additional gain that open access might provide.

The savings in time that could be achieved under an open access model are massively overstated. Houghton and Oppenheim have given us many examples of possible time savings, but they are all variants on the theme of researchers not having access to the scientific publications that they need for their research. There is nothing in the report to back up this assertion, beyond a partial reading of studies of the citation effect of open access and a calculation that most academics have access to only 50% of peer-reviewed subscription journals. There is no consideration of whether the 50% of journals – not articles – to which they do not have immediate access are at all relevant to their discipline and research.

**Estimated benefits to R&D performance**

The estimated improvement in gains to R&D investment is based on the same premise: that open access gives researchers easier access to the information they need to conduct their research and thereby improves the efficiency of that research. The overall gain is calculated at a 5% improvement, based on (i) a gain of 2% through a reduction in duplicative research; (ii) a gain of 2% through a reduction in research in blind alleys and faster publication; (iii) a gain of 2% through faster search and discovery; and (iv) a gain of 2% through more effective collaboration. The total gain of 8% is then reduced to 5% so that the authors can claim that it is ‘conservative’. Yet, not one of the claimed gains of 2% stands up to scrutiny.

If 2% of current research is duplicative, this is not because of poor access to the research literature. Academic researchers have immediate and seamless access to the vast majority of the literature that they need for their research, through direct subscriptions to journals, the Big Deal and aggregation services; and they have 100% access to the literature through indexing and abstracting services for their disciplines, discovery tools like Google Scholar and Scopus, the informal communication channels that exist within every area of research, and inter-library loan and document delivery services. The critical issue is not how much of the literature researchers have access to; it is how well they use this access.

Again, if 2% of research is down blind alleys, this is not because of poor access to the literature. It is a matter of the competence of researchers in using the information sources available to them.

With the claimed gain of 2% through easier search and discovery, we are back to the notion of citation advantage and faster research. We have already seen that the citation advantage is probably non-existent, yet the report is able to pluck another 2% gain out of the air. Finally we have the notion that the marginal additional access created by open access, if it exists at all, somehow leads to a further 2% gain through collaborative research which could not otherwise have happened.
So, the report plucks out of the air four gains to R&D efficiency, each of 2%, with flawed justifications for each of them, and then by reducing the combined value of 8% to just 5% the overall gain is said to be ‘conservative’ and a ‘plausible starting point for modelling’. In fact, the report’s authors have failed to show that there is any real gap between the access that academic researchers have today to the scientific literature that they need, and that which they might have under an open access model. And all that follows is therefore science fiction. You can build all the mathematical models you like, but if you put partial or bad data in you will get partial or bad data out.

Conclusion

Of course, it is most unlikely that an entirely funder supported producer-side OA publishing system would arise. (Houghton et al., 2009, p. 145)

Why, oh why, was this eminently sensible statement not borne in mind throughout the report? Why model a purely open access world if you accept that it is not going to come about? If you believe that some sort of mixed model will continue, with perhaps different proportions to those that pertain today, why not try to model for that? Because if you did model for that, many of the assumptions that you make about the purported savings and efficiency gains would disappear, and that would not suit the case that you want to make.

Nor does the report bother itself with how we might get from A to B, from this less than perfect world to that promised land. But the report is not really interested in relating to the real world; if it had been, it might have invited a broader participation in its development. The report is a house built on sand. Its foundations are myriad estimates and assumptions, many of which are simply unsupported by evidence.

Collaborative approaches to extending access to scholarly literature

In contrast to Houghton and Oppenheim’s ‘independent’ report, there have been several initiatives involving most or all players in the scholarly communications chain in the last year which aim to review, in an open and collaborative way and based on reliable and accurate data, the implications of the move to online-only publishing, the gaps in access to scholarly information, and how scholarly publishing might take advantage of emerging technologies and business models effectively and sustainably to extend access to the scientific literature. It might be argued that the publication by JISC of the original report, and the hostile reaction to it by publishers, galvanised the UK participants to cooperate more closely and in good faith to undertake these joint initiatives, in which case it will have facilitated valuable progress.

There is no space here to look at all these in detail, but the initiative most relevant to the Houghton and Oppenheim report is the recent ‘joint statement’, Transitions in Scholarly Communications – a Portfolio of Research Projects, issued in November 2009 by the Research Information Network, JISC, RCUK, the Wellcome Trust, Universities UK, SCONUL, RLUK, the British Library, SPARC Europe, the Publishers Association, ALPSP, STM and the Publishing Research Consortium (RIN, 2009).

It’s difficult to imagine a broader or more inclusive collaboration. The statement acknowledges that while all the players want to see access to the research literature widened, there is no consensus as yet as to how this might best be achieved:
The scholarly communications landscape has been transformed over the past few years, in the UK and across the world. Technological change has brought – and continues to bring – profound changes in the roles that researchers, funders, research institutions, publishers, aggregators, libraries and other intermediaries play in disseminating and providing access to quality-assured research outputs, in their goals and expectations, and in the services they provide and use. There are shared ambitions for significantly enhanced access, but no consensus on how best to achieve it.

The statement then sets out a portfolio of work which will lead to a better understanding of the changes that are taking place in scholarly communications and thus how new technology and business models might be exploited to best effect.

Two US initiatives with similar aims of fostering collaborative work among publishers, libraries and the scholarly community are the Chicago Collaborative (Chicago Collaborative, 2009) and the House of Representatives Scholarly Roundtable (AAU, 2009). In Europe, there are the PEER project, Publishing and the Ecology of European Research (PEER, 2009), which is well established and now in its second year, and the more recent Study of Open Access Publishing (Project SOAP, 2009). What all these projects share, apart from their collaborative nature, is a desire to ensure that any decisions which lead to fundamental changes in the current modes of scholarly communications are based on firm evidence and an equal desire that any transition from the current model to any future model is managed effectively, and that any risks that go with such a transition are mitigated.

One cannot prejudge any of these projects, but instinct suggests that it is highly likely that a more mixed model will develop, with subscription publishing, author-pays open access and repositories all playing a role. Just what kind of balance there will be between these and any other models we cannot yet say.

References


