Widening access to research information: collaborative efforts towards transitions in scholarly communications

Steven Hall

Berlin 7 Open Access Conference, Paris, 2 December 2009

Introduction

Let me start with a brief preamble about my credentials to speak on this topic. I have worked in the academic publishing industry for more than thirty years. For four or so of those thirty years I worked in the publishing of scholarly journals and books, at Blackwell Publishing. I no longer work directly in this particular sector of the commercial publishing industry and have not done so for the last eighteen months, during which I have been acting as a consultant to quite different parts of the industry. So, lest you think that I come to this meeting with my livelihood dependent on the commercial publishing of journals and monographs, I do not. The International Association of STM Publishers asked me to speak on its behalf at this conference partly because I have no particular corporate affiliation and can speak for the industry as a whole.

So I come to this topic from a certain distance – but with a deep knowledge and experience of commercial academic publishing and with a recognition of both its good and bad sides.

What I would like to do today is to review the some of the assumptions and estimates on which the *Economic Implications of Alternative Scholarly Publishing Models* is based; and then to talk about some of the initiatives which are taking place between all the various players in the scholarly communications chain to understand better the current transitions in scholarly communications and to take advantage of them to widen access to research information.

Review of the *Economic Implications of Alternative Scholarly Publishing Models*

Given the constraints of time in this presentation, I am going to focus on journals and on the economic impact of the author-pays open access model as estimated by Professor Houghton and his collaborators. That is, I am largely going to ignore the sections of the study which relate to monographs, as these are even more speculative than the sections on journals, as the study itself acknowledges in places; and I am also largely going to ignore the estimated impact of the self-archiving and publishing overlay model, as this exists in the realm of fantasy only, with no real-world model on which to base the costings or the estimates of savings and benefits.

I am also going to focus on the study of the UK. While I am aware of Professor Houghton's studies of the Netherlands, Denmark and Germany I believe that the methodology of each study is largely the same.

Let me say loud and clear at the start that publishers are not opposed to Gold open access. As the STM association said in its Brussels declaration in November 2007 (http://www.stm-assoc.org/public_affairs_brussels_declaration.php), publishers are generally agnostic about business models. Gold open access has potential as one of a series of emerging business models which could help to maximise the dissemination of research on the web in an economically

sustainable way; and a good case can be made for it. Unfortunately, it's not the case that Professor Houghton and his collaborators have made for it.

The report's authors have clearly engaged in a very wide-ranging and detailed review of the literature on scholarly publishing, and for this they must be commended. But they have chosen to ignore or, at best, to minimise the significance of any research which is at all critical of the various forms of open access while playing up every piece of research which supports it. Similarly, there are myriad assumptions and estimates about the relative costs and benefits of 'toll-access' publishing and the Green and Gold open access models, entirely in favour of open access models and many of them plain wrong.

There was no engagement whatsoever with the subscription publishing industry during the report's development. In response to criticism of this kind the JISC has claimed that '*The study was intended to be independent. Formal consultation with representative groups from any particular part of the domain would have risked this independence.*' Independent of what? How could the study ever be independent when it was commissioned from a strong advocate of open access and its 'expert review group' consisted largely of more supporters of open access?

A more open set of terms of reference and the involvement of representatives of the publishing industry would have ensured a more balanced and credible report. As it is, the study does not have a single positive thing to say about the existing publishing paradigm, one which has survived for several centuries and which has invested very heavily in the last fifteen to twenty years to facilitate the transitions enabled by new technology; nor does it have a single negative thing to say about the prospects of open access, in neither of its forms, nor any analysis of any potential downsides. It is an extraordinarily one-sided view of the topic. This is all the more to be regretted at a time when there appears to be a growing appetite for constructive collaboration between all the players in the scholarly communications chain.

Let me give you just one example of the report's bias. In section 3.2.1 entitled *Access issues and limitations* the report 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'. It supports this statement with three anecdotes – yes, anecdotes. Now, no-one would claim that there is no knowledge gap between the industrialised world and the developing world, but for the report to publish three anecdotes in support of its own case and to make no mention of the efforts that publishers and others are making to bridge that gap is simply dishonest.

Three programmes, the first dating back to 2002, provide free or very low-cost online access to more than 7,500 peer-reviewed scientific journals, plus many books, indexes and databases, to researchers at 4,500 institutions in the developing world. These are, of course, the HINARI, AGORA and OARE projects under the umbrella of Research4Life, a public-private partnership between the WHO, FAO, UNEP, Cornell and Yale Universities and more than 130 science publishers, along with their technology partner Microsoft. Institutions in more than 70 countries have entirely free access, while institutions in another 36 countries have very low-priced access, with all income reinvested in local training initiatives.

In support of its open access case the report claims that while institutions in the developing world have little or no access to subscription content, they do make use of open access content, noting, for

example, that the Australian ARROW Discovery Service had two hits from Benin in 2007 and 21 from Zambia. Well, under Research4Life 21 institutions in Benin have completely free access to 6,562 journals in HINARI, 18 have access to 1,278 journals in AGORA and 17 have access to 2,963 journals in OARE. There is a similar picture in Zambia. In total, in 2008 there were more than 6.5M PDF article downloads from HINARI, AGORA and OARE by researchers in the developing world.

These three projects are long established and well known. So why provide three anecdotal pieces of evidence for poor access to scholarly journals in the developing world and ignore firm and published evidence of improving and completely free access? Either it didn't suit the case that the authors wanted to make; or they missed it because they were poorly advised.

Let's move on to look at the claims that the study makes for cost savings and additional economic benefits to be gained from a transition to open access. Let's start with the estimates for cost savings from publishing.

Publishing cost savings

One of the difficulties in responding to the study 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.

In places the report acknowledges these difficulties:

Perhaps, greater differences lie in the switch from print and dual-mode publishing to an e-only model – although disentangling the essential cost differences of alternative publishing models from 'print economics' and the economics of digital delivery (Cockerill 2006 and OECD 2005) is by no means easy as OA publishing and self-archiving entail online delivery while toll access can be based on online, print or dual-mode delivery.
Houghton et al, 2009, p. 79

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

Frankly, I don't think it's all that difficult to look at the economic impacts of a move from the current mixed model to a purely online model. All publishers are modelling what it will mean as their businesses move from today's mix to a largely online future. It's the report's methodology that is wrong. It would have been perfectly straightforward and more transparent to have done this work in two separate stages, as the RIN/CEPA study did, first looking at the impact of a move to largely online-only (with print-on-demand as an adjunct to it), 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.

Table 4.14: Potential cost implications of OA publishing for UK higher education (GBP, circa 2007)

	Estimate
Implied total cost (National)	171,700,000
Implied total cost for Higher Education	147,500,000
Implied saving if the World adopted 'author-pays'	-34,800,000
Implied saving if UK HEIs adopted 'author-pays'	-139,100,000
SCONUL Library savings (with subscription cancellations)	156,400,000
Potential net saving for UK Higher Education Institutions	8,800,000

Note: All costs are expressed in 2007 UK pounds and, where necessary, have been converted to pounds using OECD published annual average exchange rates and adjusted to 2007 using the UK consumer price index published by the National Statistical Office.
Source: EI-ASPM model: Authors' analysis.
Houghton et al, 2009, p. 181

In this table - *Potential cost implications of OA publishing*... – the estimated savings to libraries from the move from the current mix of print and online are included alongside the benefit of a move to open access. This is misleading.

Making open access publications available internally *demands less effort on the part of research and institutional libraries (Figure A4223). Core activities can be limited to cataloguing and/or providing links to OA content, and providing reader/user support. OA publishing and self-archiving models do not require such activities as subscription processing, establishing and operating an authentication system or checkout and handling. As result, there is potential for substantial cost savings.*
Houghton et al, 2009, p.93

And again. Savings on checkout and handling have nothing whatsoever to do with open access publishing; they are all to do with the move from print to online. I'll come back to the issue of whether authentication systems are required.

- *Providing the foundation for the integration of all the elements of the outputs of e-research (i.e. institutional repositories catering for all sorts of digital objects).*
- *Providing the foundation for long-term preservation (e.g. distributed preservation through institutional or subject repositories).*
- *Enabling the development of new mechanisms for collaborative research and underpinning new forms of research.*
- *Reducing the operational costs of libraries through the switch to OA 'e-only' content (e.g. less space, shelving, handling, etc.).*

Houghton et al, 2009, p. 133

And once more the savings and benefits in the move to e-only are conflated with the savings in a move to open access.

I could go on. There are many other instances in this document where the benefits of a move to e-only are mixed up with the estimated benefits of a move to open access.

Publishers would generally be happy to move from the current mixed model to an online-only model. What they can't do, though they have lobbied for it, is to 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.

Let's look now at some of the specific publishing cost savings that the report identifies.

First, there are payment handling costs.

Author-side payments processing: *For the author-side fee model of OA publishing the cost of processing author-side payments is estimated at £20 per article, based on CEPA (2008) estimates.*
Houghton et al, 2009, p. 155

Sales administration and online user management: *For the subscription model, sales administration and online user management are estimated to each cost £10 per subscriber (CEPA 2008). Costs per article vary with the number of subscribers, but at an average of 1,200 subscribers this would be equivalent to around £100 per article published.*
Houghton et al, 2009, p. 155

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. What it 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. It's the wrong comparison to make. 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 higher education institutions. That's just the cost to publishers. Then there is the cost to authors in terms of their time, and the cost to their institutions or funders in terms of administering author payments, which are not included in the cost estimates. The assumptions here are flawed; and, in passing, why does the report reduce the RIN/CEPA estimate of £30 on the publisher side to £20?

Then there's rights management.

Rights management: *Based on CEPA (2008), we estimate rights management costs at around £50 per article for toll access copyright-based publishing, and £10 per article for OA publishing – with standard licensing agreements progressively replacing individual copyright assignment.*
Houghton et al, 2009, p. 155

On what basis do publishers save 80% of their rights management costs? The fact is that today every corresponding author signs some form of agreement with the publisher, the vast majority with no problem whatsoever. Under an open access model every corresponding author will be required to sign some form of agreement. The volume doesn't change. There is no justification at all for the claim of an 80% cost saving. This is another flawed assumption.

Then there's marketing, where there's an equally unsupported claim for a cost saving of 67%.

Marketing: *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

I have no idea what sources the authors drew on, but they clearly weren't 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 in 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.

Professor Houghton and his collaborators seem to think that in an open access world such marketing is not going to be necessary. If we assume that there will be competition between open access publishers, and ignore the recent suggestion by one open access advocate that for-profit publishers should be 'eliminated', then marketing will be necessary. They also ignore the large sums spent on marketing by open access publishers today, including, for example, primetime television advertising.

The predicted savings are a fantasy; the assumptions behind them are deeply flawed.

Similar claims are made for savings on online hosting, with savings of 50% estimated.

Publisher – distribution costs could be reduced in OA publishing through the use of OA repositories (e.g. PubMed Central) or hosting services (e.g.HighWire) instead of using proprietary in-house systems for distribution.
Houghton et al, 2009, p.81

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; and it possibly stems from a misunderstanding of the CEPA data. Publishers will want to differentiate their services from those of their competitors, through value-added services to users, authors, reviewers, etc. They will not want simply to deposit articles on a platform like PubMed Central. The reference to Highwire here also shows another misunderstanding of how publishers deliver their online journals today. They don't all use 'proprietary' systems. A good number use platforms like HighWire – OUP and Sage, for example – and others use third-party services like Metapress – Springer, for example. They choose their platform on the basis of functionality, cost and other similar measures. They don't choose HighWire because it is inexpensive; as a highly functional platform, it isn't.

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 be bound

to be additional investment in systems to understand usage better, which could no longer simply be captured at institutional level.

Next we have customer service and the user help desk, with more estimated savings of 80%.

Customer service/helpdesk: *Following CEPA (2008) we estimate the cost of operating customer service/helpdesk at £50 per article for the subscription model, and £10 per article for the OA publishing model – with no subscriber access problems to deal within the latter.*
Houghton et al, 2009, p.155

The report assumes that in this brave new world publishers will simply throw their content over the wall and leave users to get on with it. There may be some small saving were authentication no longer to be required, but again this is a small part of costs. Or will all platforms function perfectly all the time for every user in the open access world? Once again, the assumptions about how online publishing works today, and in this case about how it might work in the future, are flawed.

Next up there's management and investment margins.

Management and investment: *Following CEPA (2008) we allow a management and investment margin of 20% for management and investment. This accords with industry consultation. For the OA publishing model we allow a margin of 15% due to reduced overheads in relation to such things as pricing, proprietary hosting systems, legal and licensing, reduced investment as author-fees materialise immediately, etc.*
Houghton et al, 2009, p.157

and

The development of more sustainable business models, as subscription revenue is becoming difficult to sustain in the face of declining subscriptions (e.g. moving to 'author-pays' OA publishing would make revenue more predictable and stable as it scales more easily to research output than have library budgets, growing with research expenditure and providing a revenue stream that is growing) (e.g. Waltham 2006b). This reduced level of risk should, over time, be reflected in a reduced user cost of capital for OA publishers.
Houghton et al, 2009, p.131

We can take issue with two more suggestions here. The first is that revenue would be more predictable under an open access model. One of the big attractions for those who invest in the scholarly publishing industry today is the predictability and stability of subscription income; that's, frankly, one of the things that the opponents of commercial scholarly publishing least like about it. The second is the idea that investment will be reduced as 'author fees materialise immediately'. Can the report's authors explain to me in what way the payment of author fees shortly in advance of publication compares 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 proposed open access model are, in fact, negative, not positive. Yet again, the estimated 25% saving here is a fantasy based on a profound misunderstanding of how scholarly publishing works.

The last of the estimated savings is in the operating margins under open access.

Surplus/profit: *Operating margins are relatively high in scholarly publishing, and we allow 20% for toll access publishing and for the OA publishing model we allow a margin of 15% due to lower risk and reduced cost of capital.*
Houghton et al, 2009, p.157

On what basis are we to accept that commercial open access publishers – and the assumption here is that commercial publishers would continue to provide publishing services – will accept a lower profit margin than they enjoy today? We have already seen that there is no lower risk. This is simply wishful thinking.

Here's another piece of wishful thinking on publishing costs, though at least no economic benefit seems to have been attached to it.

... OA journals are more visible and more likely to attract submissions and advertising revenue, as well as readers and citations, thereby increasing potential revenue growth opportunities
Houghton et al, 2009, p.131

Where is the evidence to support this, other than in a very partial reading of the literature on the impact of open access on citations, which I will come back to later? There is, simply, no evidence whatsoever to support the assertion that open access journals are more likely to attract more submissions or more advertising revenue.

In passing, it's worth noting that a good part of the analysis of current costs in the report appears to be based on the research published in 1999 by Halliday and Oppenheim on the *Economic Models of the Digital Library*. 1999 is a very long time ago in digital publishing, only two or three years into the online publishing of journals and well before the advent of the Big Deal, the ebook, Google Scholar and Scopus, and so on. The research is, quite simply, out-of-date.

Finally, before we leave the topic of the estimated savings on publishing costs, let's look at the bigger picture.

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 we 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 SCONUL figures, of £113M on serials and £56M on books. So under the open access model the UK would be saving 71% of its current expenditure on journals (without any subscription cancellations) and would be saving

£38M more than it is actually spending on books. Under the self-archiving model, the UK would be saving £3M more than it is currently spending on journals (again without any cancellations) and £46M more than it is spending on books. And this is all before any author-side payment costs.

Now I'm beginning to understand why the global economy is in such a mess. No, let me not blame our current mess on economists. Let me praise them for finding the solution to the economic crisis. Let's simply apply the economics of the self-archiving and overlay model to the economy as a whole – I'm sure there's a mathematical model for it – and, hey presto, the economic crisis is solved.

This is, of course, the problem of developing a theoretical model with little or no relation to the real world. It's also the problem of trying to build a model up from the cost of a single article or book. You make a few errors in your assumptions and estimates, gross the whole thing up and your numbers simply don't match anything that anyone operating in that real world would recognise. It's also the problem of trying to look at the cost of UK research output against UK spending on global research output. You can't 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

I'd like to move on now to look briefly at the estimates for savings in library costs and research performance.

The study estimates potential savings for UK higher education libraries of £34M from a move from e-only publishing and an additional £11M from the move from subscription publishing to open access publishing. In the response to the Houghton report from the Publishers Association, ALPSP and STM, it was pointed out that to realise the savings of £11M more than 200 librarian jobs would have to be lost. The JISC 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*.' Well, jobs are being lost in libraries right now, because of the recession, and it's hard to see the Government not wanting to realise these savings if the opportunity came along. You cannot have your cake and eat it. You either realise real savings or you don't.

Furthermore, the study itself is highly equivocal about the role that libraries play in enabling easy access to the right content for the users they serve.

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

One the one hand, the report suggests massive savings in research performance through the easier discoverability of relevant content in an open access world (which I will address shortly); 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 can't have it both ways. Libraries play an essential role in providing access to the content that their specific user communities need. Increasingly the

librarian's role is about enabling discovery of content rather than physically collecting it. The study simply doesn't get this.

The report assumes that the same level of savings could be achieved in special libraries as in higher education libraries. I'd suggest that this shows a lack of understanding of special libraries which, again, with some consultation could have been avoided.

Assuming similar library handling times and staff costs to those in SCONUL libraries, would suggest annual special library journal handling costs of £22 million if all subscriptions were print and £6 million if they were all electronic, and perhaps £4 million if they were all OA e-only. Hence, with around 15% of the number of SCONUL research library subscriptions, special library handling costs are likely to be around 15% of those in SCONUL libraries, as are the potential cost savings.
Houghton et al, 2009, p. 179

This would likely be a wrong assumption. The biggest users by far of STM journals are pharmaceutical companies. Several such companies will be amongst the top 100 customers of most major STM publishers, with some likely in their top ten customers. Most have moved almost entirely electronic; they have already realised their savings from a move to e-only. GlaxoSmithKline, for example, operates a global virtual library and buys nothing in print any longer. GSK also manages its global virtual library, for a staff of many thousands, with just twelve people. The savings to be realised in these libraries, in terms of handling costs, will be minimal.

One final point on libraries relates to the following misleading claim:

To date, relatively little attention has been paid to the long-term preservation of electronic and/or OA content, although it is clear that preservation costs are likely to be substantial. However, those costs are unlikely to vary greatly between publishing model (e.g. between toll and open access).
Houghton et al, 2009, p. 94

Again, one wonders at the lack of knowledge of what has been happening in electronic publishing over the last five years or so. There have been numerous and substantial initiatives by a number of organisations over that time, including a joint project between the British Library and the JISC and other work by the BL and JISC independently, both with the active collaboration of journal publishers; work by the Koninklijke Bibliotheek in the Netherlands, which has been operating an archive for electronic journals for several years, again with the participation of many publishers; by JSTOR and its offshoot Portico, with which most large STM publishers are depositing their electronic journals; and by the collaborative LOCKSS/CLOCKSS project between major academic libraries and STM publishers.

As to whether the costs of preservation will vary greatly between subscription publishing and open access publishing, then much would depend on how the open access model developed. It is one thing for an archive to ingest very large amounts of electronic content in standard and consistent formats from the major STM publishers; it would be quite another for such an archive to ingest content from many smaller publishers, never mind from individual authors, in huge varieties of formats and standards. If this study had attempted to be even-handed, it would have recognised some of the potential problems of open access publishing, and more especially of self-archiving, and quantified some of the potential downsides. Instead, it chose to look at the open access world through a pair of rose-coloured spectacles, this promised land, this garden of Eden.

Author fees

One of the largest single figures in the study is that for the cost of author-side payments in relation to UK published output: £170M, or £150M in higher education.

There is an enormous range of author-fees, but based on a 'straw poll' of eight major OA publishers we find fees being charged ranging from around £800 to £1,600, with a consensus in the sample and literature of around £1,500. It has also been noted that author fees are coalescing around USD 3,000 (around £1,500 at 2007 average annual exchange rates) (Kiley 2007). These reported fees triangulate well with the OA publisher cost estimates outlined above, which were derived independently from a bottom-up costing of publisher activities, in which OA e-only costs per article were estimated to average £1,448. At 2007 levels of article publishing by UK researchers, author-side fees (or producer-side fees) would have cost the UK around £170 million nationally for all journal articles published, of which around £150 million would have related to higher education.
Houghton et al, 2009, p. 180

Moreover, with author-fees charged converging at around USD 3,000 or £1,500 (at 2007 annual average exchange rates), our estimate of £1,525 per article for e-only OA publishing would seem reasonable given that it includes full commercial margins.
Houghton et al, 2009, p. 157

There is very little more in the study on the level of these fees. The subject required much more than the cursory analysis it was given.

Let's take a look at the charges which Gold open access publishers are levying. Most commercial publishers are charging in the region of $3,000 (considerably more than £1,500 at current exchange rates and likely to stay that way); most have also said that their current prices are experimental and may change in the future. Some impose colour charges on top of the author fee, which the study appears to ignore. BMC's initial article processing cost was $525; it's now between £925 and £1,025, or $1,850 and $2,051 using the same exchange rate used in the study. So BMC's charges have increased by between 3.5x and 3.9x in a very short period of years. PLoS has likewise increased its author charges very considerably since its launch and now charges $2,900 for PLoS Biology and Medicine, $2,250 for three more journals and $1,350 for PLoSOne with its much lighter peer review; and we know that even at these prices, more or less in line with what commercial publishers are charging, it is not yet covering its costs and it is almost certainly subsidising its other journals with the income from PLoSOne which doesn't use the same business model as other scholarly journals. And even if PLoS does finally break even on an annual basis in 2010 or 2011, then it will still take many years before it has broken even on a cumulative basis; and that's with close to $12M of grants and donations behind it between 2004 and 2008. Note too that PLoS is carrying none of the costs of dual-mode publishing and in theory doesn't need to make a profit, yet it is still struggling to break even.

So the study is simply wrong in presenting £1,500 as a reliable and sustainable author publishing fee, and to suggest that this figure includes a full commercial margin. It's more wishful thinking, based on a flawed understanding of the economics of scholarly publishing.

If the appropriate fee were closer to $4,000, which might still be conservative, then those UK costs would increase from £170M to £226M at the exchange rate used in the study. If we then applied

the current exchange rate instead of one from two years ago, those costs would further increase to £272M. But what's £100M between friends?

So far we have been dealing largely with figures that could be known and extrapolated from, even if the study failed to get a good number of them right. But as we move from looking at publishing costs to looking at the potential savings in research performance, and then further to the estimated increased returns to R&D, we move much more from the non-fiction shelves towards those for (science) fiction.

Research performance savings and increased returns to R&D

I am going to address the research performance savings and the increased returns to R&D together because they are both based on the same assumptions and estimates of the benefits of improved access to research information.

The study suggests 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 study is misleading on current access provisions).

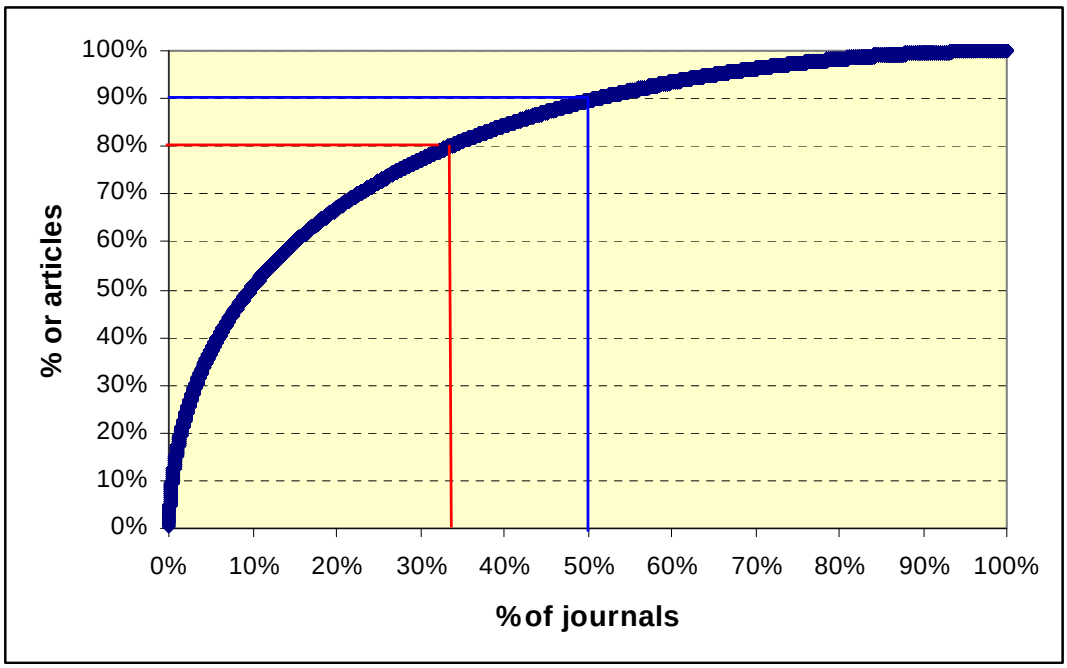
This is all largely predicated first on researchers and others having access to less than 50% of the research output that they need:

By no means all serials are journals and there are likely to be duplicate subscriptions within institutions, but even if they were all journals, all peer reviewed and there were no duplicate subscriptions the mean of SCONUL library subscriptions would represent around 39% of titles and the median 27% – suggesting that perhaps 60% to 70% of possible titles are not being made available in this way.

Björk et al. (2008) suggested that around 11.3% of articles are green OA and a further 3.5% delayed OA, so perhaps 15% of toll access content is already available OA and should not be included in estimates of the accessibility differences between toll and OA publishing and self-archiving models. 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

Let's look at this first and then the other 'evidence' for lack of access under subscription publishing, the so-called citation advantage.

The report assumes that 50% of journals equal 50% of articles. This is simply wrong. 33% of journals account for 80% of all articles published and 50% of journals account for 90% of all articles published, as this table on journal article distribution (data source: Scopus) shows.



Now let's look at the citation advantage where the report seems to me to have been very partial in its selection and review of the relevant literature.

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

In 2007, in an article in the Journal of Infometrics, Craig and others in a critical review of the literature up to that point concluded that '*the so called OA citation advantage has little to do with OA status. Citation differences can be attributed to other factors that studies claiming advantage had not adequately accounted for in their studies. At the time of our review, the most rigorous study that addressed this issue and showed that OA advantage did not exist was that of Moed ... who is a well respected and recognised expert in the field of bibliometrics. Up to that point most of the core bibliometrics community had not yet engaged in this issue. Those claiming OA advantage were generally outside of the core bibliometrics community. Since the review a number of other papers ... have concluded the same*.' (Craig et al, Do open access articles have greater citation impact? A critical review of the literature, Journal of Infometrics, 2007).

In the first randomised controlled trial in 2008, Davis and others concluded that *'No evidence was found of a citation advantage for open access articles in the first year after publication'* and *'The citation advantage from open access reported widely in the literature may be an artefact of other causes.'* (Davis et al, Open access publishing, article downloads and citations: randomised controlled trial, BMJ, 2008).

In a study of the Astrophysical Journal in 2007 Kuntz and Henneken stated that '*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.'* (Kuntz and Henneken, Open Access does not increase citations for research articles from The Astrophysical Journal, ArXiv, 2007).

And in an article in the Journal of the American Society for Information Science and Technology in 2007, Moed concluded that '*The analysis provided evidence of a strong quality bias and early view effect. Correcting for these effects, there is in a sample of six condensed matter physics journals studied in detail no sign of a general open access advantage of papers deposited in ArXiv.*' (Moed, The Effect of "Open Access" on Citation Impact: An Analysis of ArXiv's Condensed Matter Section, Journal of the American Society for Information Science and Technology, 2007).

Not one of these four articles is even listed in the bibliography to the study, never mind reviewed for it. This is a partial reading of the literature.

In 2009 two further studies of the so-called citation impact have been published. In an article in the Journal of Infometrics Frandsen concluded that '*the result of this study does not provide evidence of an open access advantage for working papers in economics* ' (Frandsen, The effects of open access on un-published documents: A case study of economics working papers, Journal of Infometrics, 2009) and Lansingh and Carter came to the conclusion that **'***it appears that open access in some ophthalmology journals, whether such journals are completely open access or subscribe to a mixed model, has not influenced the manner in which ophthalmology research has been cited*' in a study published in Ophthalmology in August 2009 (Lansingh and Carter, Does Open Access in Ophthalmology Affect How Articles Are Subsequently Cited in Research?, Ophthalmology, 2009).

Once again we have to conclude that the authors' analysis and interpretation of the published research has been flawed. At best, the jury is still out on the effect of open access on citations, though the most recent research appears to be showing no effect at all, across a range of disciplines.

Let's look then at the supposed savings in research performance. Here's a selection:

As noted, the main areas of cost impact of alternative publishing models relate to accessibility and usability – content costs, search, discovery and retrieval time, and permissions to use and re-use as required. These include:

- *Possible cost impacts on search and discovery (e.g. proprietary access systems used to control toll access involving use of multiple interfaces and imposing learning and switching costs on users, etc.). Halliday and Oppenheim (1999, p17) reported that the variety of systems proved frustrating to users.*

- *Possible differences in the timeliness of retrieval (e.g. delays caused by authentication slowing download, payment barriers and transaction time, document delivery, etc.).*

- *Potential for OA to reduce or eliminate the delays, costs and frustration caused by access barriers when trying to follow hyperlinks*

- *Self-archiving allows user choice – immediate free access to the pre- or post-print versus access to the value-added version in a journal (Green OA).*

- *Possible differences in the availability and usefulness of alerting services (e.g. difficulty of replicating the thematic bundling of material achieved by journals directly from OA institutional repositories, although subject repositories can provide thematic alerting).*

Houghton et al, 2009, p. 116

In the first example we are once again quoted Halliday and Oppenheim, from way back in 1999. The fact is, the world of online delivery has moved on since then. I haven't heard a librarian complain about the difficulty of managing multiple journal interfaces for many years. Why? Because the interfaces to the various journal platforms have become quite similar over the years and above all because the vast majority of researchers use the publisher platform only to retrieve an article; they do their searching in Google Scholar (50% of a journal platform's hits can come from Google Scholar alone) or Scopus or an abstracting and indexing service or a library's federated search tool or they simply follow a link from another article or from a publisher's alert; and so on. This is simply not an issue and there is no open access saving here. And, furthermore, who is to say that all open access content platforms are going to be the same?

On the second point, where is the delay caused by authentication? Users authenticate on their library's own platform, not on every publisher's site. And though it's suggested elsewhere in the report that libraries will save through not using authentication systems, that could only be true if every electronic publication accessible through a library's online portal was open access. Are all reference works going to be open access too? All databases? Every textbook? Every online newspaper and magazine? No, of course they are not. Libraries will still need authentication tools to ensure that only their users have access to licensed content provided through the library.

On the third point, linking is already more or less seamless, given that researchers already have access to the vast majority of the materials they might require. Again, the open access saving is negligible if not non-existent.

On the fourth point, publishers are already beginning to make pre-prints of articles available online, replacing them with the final published version when it becomes available. This is not a benefit of open access, simply one of electronic publishing.

And on the fifth point – oh my goodness, a possible downside of open access, though as far as I can see it's not quantified anywhere as a possible additional cost or loss of efficiency.

The list of these supposed savings in time seems to go on and on, and they are then all repeated later in the study in the list of supposed benefits to R&D of open access, though they are supplemented there by some wilder claims still

The fact is, researchers today have immediate access to the vast majority of the scientific articles that they could need for their research. This is thanks to two things, the first being years of fine-tuning of collection development by librarians to ensure that their users have access to the core journals for their disciplines; and 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 in the UK over the last ten years.

In the companion report to a study entitled *Access to professional and academic information in the UK* published in August 2009 by the Publishing Research Consortium, 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 doesn't need the vast majority of medical journals; the LSE doesn't need journals in high-energy physics. And so on.

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 I'd say that the Big Deal would be worth a gain of at least 10% in research efficiency and probably more than £1B to the UK economy. Wow. For an investment of what – less than 10% on top of UK higher education's previous spending on journals – the JISC should have the exact figures from its NESLi negotiations – probably under £10M – we have a hundredfold gain. But of course publishers don't make that claim. However, what we do have here is the real world. The Big Deal exists and presumably its impact on research performance and efficiency can be measured. If Professor Houghton and his colleagues 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. Professor Houghton and his colleagues may have given us many examples of those possible time savings, but they are more or less 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 study 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 less than 50% of peer-reviewed subscription journals, without bothering to look at whether the 50% of journals – not articles, as I pointed out earlier - they don't have access to are at all relevant to their discipline and research.

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 the following:

Scenario: Less risk of duplicative research being done through greater access and more complete and faster publication

If just 2% of total research time were spent performing duplicative research that could have been avoided if researchers had had more complete and immediate access to the findings of others, then the annual 'saving' would have been more than £465 million nationally in the UK from Gross Expenditure on R&D circa 2006, of which the 'saving' in higher education expenditure on R&D would have been around £120 million – equivalent to around 8.2 million and 2.2 million researcher hours, respectively. With returns to publicly funded R&D at a conservative 20%, the implied lost annual returns (i.e. from the same amount of research expenditure that was not duplicative) would have been around £93 million nationally and £24 million from higher education. This implies a possible overall loss to the UK of up to £555 million per annum, of which £145 million could have been realised from higher education.

Houghton et al, 2009, p. 203

If 2% of current research is duplicative then it's not down to a lack of access to the research literature. As I have said before, 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 and, of course, the informal communication channels that exist within every area of research. The critical issue isn't how much of the literature researchers have access to; it's how well they use that access.

Scenario: Less risk of pursuing blind alleys through greater access and more complete and faster publication

If a similar 2% of research time is spent pursuing blind alleys could have been avoided if researchers had more complete and immediate access to the findings of others, then the same impacts and savings could be realised (i.e. £555 million nationally per annum, of which £145 million from higher education).

Houghton et al, 2009, p. 203

Again, if 2% of research is down blind alleys it once again isn't down to a lack of access to the literature. It's again down to the competence of researchers in using the information sources available to them; or it's going to happen anyway because comparable research has simply never been done.

Scenario: Enhanced and more immediate access speeds up the research and discovery process

It has been suggested that more open access not only increases citations but leads to earlier citation (EPS et al. 2006; Schwarz and Kennicutt 2004; Brody and Harnad 2005), suggesting that the research and discovery process may be accelerated. This implies that less research time would be required for a given output/outcome or that more research could be done for the same expenditure (cost and/or time). If this increased the returns to R&D spending by 2%, then it would have been worth around £465 million per annum nationally circa 2007, and £120 million for higher education alone – equivalent to 8.2 million and 2.2 million researchers hours, respectively.

Houghton et al, 2009, p. 204

Here we are back to the notion of citation advantage and faster research. We have already seen that the citation advantage is quite possibly non-existent, yet the study is able to pluck another 2% gain out of the air.

Scenario: Collaborative research made possible by near universal access to the entire body of research publications brings higher returns to R&D
It is widely held that there are advantages to collaborative research and greater use of the findings of collaborative work (Katz and Hicks 1997; Katz and Martin 1997; Walsh and Maloney 2001). If collaboration increased the returns to R&D by a similar 2%, then it too would have been worth around £465 million per annum nationally circa 2007, and £120 million for higher education alone – equivalent to 8.2 million and 2.2 million researchers hours, respectively.
Houghton et al, 2009, p. 204

And 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 somehow couldn't otherwise have happened.

So the report plucks out of the air four gains to R&D efficiency, each of 2%, with no justification for any of them beyond some wishful thinking – 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'.

The fact is, the report's authors have failed to show that there is any real gap between the access that 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, as I said earlier, science fiction. You can build all the mathematical models that you want, but if you put partial or bad data in you'll get partial or bad data out. This is put much more elegantly than I could put it in a joke, which I'll read to you now to lighten what has been a somewhat negative half-hour:

A mathematician, an accountant and an economist apply for the same job.

The interviewer calls in the mathematician and asks "What do two plus two equal?" The mathematician replies "Four." The interviewer asks "Four, exactly?" The mathematician looks at the interviewer incredulously and says "Yes, four, exactly."

Then the interviewer calls in the accountant and asks the same question "What do two plus two equal?" The accountant says "On average, four - give or take ten percent, but on average, four."

Then the interviewer calls in the economist and poses the same question "What do two plus two equal?" The economist gets up, locks the door, closes the blind, sits down next to the interviewer and says "What do you want it to equal?"

Finally, on the topic of this study of the UK, I'd like to draw your attention to one important general statement in the study:

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, wasn't this eminently sensible statement borne in mind throughout the study? 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 wouldn't suit the case that your sponsors want to make.

Nor does the study bother itself with how we might get from A to B, from this less than perfect world we inhabit to that promised land. But it's not really interested in relating the study to the real world, is it? If it had been, it might have invited a broader participation in its development.

The study is a house built on sand. Its foundations are myriad estimates and assumptions, many of which are simply unsupported by evidence or simply dreamed up by the authors.

So let me move on, and back to the real world, to look at some new and recent initiatives which involve most or all of the players in the scholarly communication chain. We have already seen from the Research4Life example that STM publishers are working effectively with other interested parties – in this case, NGOs, universities and a technology partner – to make a very real difference to access to scientific information in the developing world. Let's look at some other initiatives which aim to widen access to scholarly information but which are grounded in reality and collaboration.

The one perhaps most relevant to Professor Houghton's study is the recent 'joint statement' on *Transitions in scholarly communications – a portfolio of research projects* issued in November 2009 by – and it's a long list – the Research Information Network, the JISC, RCUK, the Wellcome Trust, Universities UK, SCONUL, RLUK, the British Library, SPARC Europe, the Publishers Association, ALPSP, the International Association of STM Publishers and the Publishing Research Consortium – so a broad coalition of research funders, research libraries, higher education bodies and publishers (http://www.rin.ac.uk/news/new-projects-transititions-scholarly-communications-launched).

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 that 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 used to best effect. There are four strands to this work, though others may be added in the future:

***Transitions to e-only publication**, which will investigate the barriers – from the perspectives of libraries, publishers and users – to moving to e-only publishing of scholarly journals, and ways in which those barriers might be overcome;*

***Gaps in access**, which will investigate the extent to which journal articles and other research outputs are available, or not, to different parts of the research and other communities which could make use*

of them; and to identify priorities in seeking to fill gaps in access, barriers to filling them, and actions that might be taken to that end;

***Dynamics of improving access to research papers**, which will provide evidence for a better understanding of the dynamics of the transitions needed to reach a selection of plausible end-points, and the costs, benefits, opportunities and risks that this entails. Transition is understood to relate to changes in practice, business models and organisational culture within the relevant constituencies, and any new entrants, over defined timeframes. The end-points, to be defined in advance of the project, will be associated with four broad models: open access journals (gold OA); open access repositories (green OA); extensions to licensing; and transactional solutions. The project will be founded on a comparative description of the transitions that (i) are taking place now, and (ii) would need to take place over the next five years, in order to reach each of the selected end-points. There will also be an analysis of the drivers and mechanisms underlying these transitions, and associated costs and benefits (both cash and non-cash);*

***Futures for scholarly communications**, which will seek to develop a series of challenging scenarios for scholarly communications in ten years' time, bearing in mind current trends and underlying drivers in user cultures, needs and expectations; and likely – and more speculative - developments in technologies and services. Through the process of developing the scenarios we shall promote constructive dialogue between stakeholders, and seek to establish priorities for action.*

Now I understand that the fine detail of some of these projects is still being discussed but you will see that overall the participants are taking a rigorous approach to them. Note that, unlike in Professor Houghton's study, they will seek to understand all the implications of moving to e-only publication separate from any analysis of potential new business models. This seems eminently sensible to me. They will also seek to identify exactly what gaps there are in to access to research information, rather than trying to guess at what they might be; understanding this is critical to determining how best such access gaps might be filled.

Another initiative with similar aims of fostering collaborative work between publishers, libraries and scholarship is the Chicago Collaborative (http://www.chicago-collaborative.org/), established in 2008 and which defines its purpose as follows:

The Chicago Collaborative was created from a conviction that we are at a pivotal moment in the history of scholarly communication. It's 1453 and we're in a workshop in Mainz trying to do a better job of getting this wine press to push these metal bits evenly onto this sheet of paper. We're in the coffeehouses of London in 1664 trying to work out a scheme for gathering and reproducing reports from the greatest scientists of the day so that we can send them all across the continent in serial issues. And it's the early 21st century, and in libraries and editorial offices and publishers meetings across the globe, we are trying to figure out how to use the internet and the web and this startling array of new digital tools to once again push scholarly communication in a new direction.

The members of the Chicago Collaborative believe that collaboration is essential to the development of a scholarly communication system that serves the best interests of the entire scholarly community. The future of scholarly communication will, we believe, be determined not only by the opportunities

and challenges before us, but by the way we approach them. While the opportunities before us today far exceed the imaginings of those visionaries from past eras, making the most of these opportunities will require the active, constructive engagement of not just some but all stakeholders.

The founding members of the Chicago Collaborative believe that success in this crucial endeavor will require a relationship of trust and a spirit of collaboration among all stakeholders. Without this essential engagement, we will fail to meet our responsibilities to society.

It is in order to create an environment where these relationships can flourish, and the grand challenges for scholarly communication of our age can be confronted, that we have founded the Chicago Collaborative.

Chicago Collaborative Founding Members:

- *Association of Academic Health Sciences Libraries*
- *Association of American Medical Colleges, Council of Academic Societies*
- *Association of American Publishers, Professional and Scholarly Publishing Division*
- *Association of Learned and Professional Society Publishers*
- *Federation of American Societies for Experimental Biology/DC Principles*
- *International Association of Science, Technical & Medical Publishers*
- *International Committee of Medical Journal Editors*
- *Society for Scholarly Publishing*

Third, and again in North America, there's the House of Representatives Scholarly Roundtable whose aims are summarised in the following status report dated 29 October 2009

The Committee on Science and Technology of the U.S. House of Representatives, which has oversight of the federal civilian R&D enterprise, has a strong interest in wide dissemination of the results of federally funded research, making these results available to other researchers and the broader public. In early summer 2009, the Committee convened a Scholarly Publishing Roundtable representing key stakeholders, who were charged with seeking consensus on feasible and effective ways to expand access to and preservation of federally funded research information.

Further information on the Roundtable can be found on the web site of the Association of American Universities at http://www.aau.edu/policy/scholarly.aspx?id=6894 .

As with the 'joint statement' participants and the Chicago Collaborative, the aim of this group is for all stakeholders to work together to identify 'feasible and effective' ways of expanding access to research information.

Finally, in Europe there's the PEER project, Publishing and the Ecology of European Research (http://www.peerproject.eu/), which is well established and now in its second year. Again, this is a collaborative project between publishers, repositories and researchers to assess the effects of large-scale systematic deposit of final peer-reviewed manuscripts on reader access, author visibility, and journal viability, as well as on the broader ecology of European research.

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 different future model is managed effectively and that any risks that go with such a transition are mitigated.

I cannot prejudge any of these projects, but my instinct tells me that is highly likely – and I believe that Professor Houghton and I may be able to agree on this - that a more mixed model will develop, with subscription publishing, Gold 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.

Let me reiterate what I said earlier at the beginning of this talk: publishers see Gold open access as a feasible business model which may well make a significant contribution to the wider dissemination of research output. They are more than willing to embrace the model and make it work. They will do so, however, on the basis of a good understanding of its economic implications, and not on the basis of the kind of flawed assumptions and misinterpretations that are all too evident in the Houghton report.