

# Will the parasite kill the host? Are institutional repositories a fact of life – and does it matter?

Based on a presentation given at the 30th UKSG Conference, Warwick, April 2007. Note this draws heavily on other articles by the author.

Despite an apparent lack of enthusiasm among academics themselves, institutional repositories seem set to grow. Two studies have highlighted the possible damage which could be caused to journal subscriptions by widespread self-archiving. If journals were damaged financially, the scholarly community would lose some functions which it appears to value very highly: management of peer review; editing; selecting and collecting content into a convenient package. It would also suffer indirectly, if learned societies were no longer able to give the same support to their disciplines. However, publishers cannot afford simply to oppose these developments; rather, they need to work with the scholarly community to identify those functions which are of greatest importance to the community in the digital era, and then to work out how to deliver and market these.



**SALLY MORRIS**

Partner  
Morris Associates

## Introduction

There seems to be little evidence that academics actually want institutional repositories. Recent studies<sup>1,2,3,4</sup> suggest that academics are not terribly aware of their institution's repository if it has one, and even when they are they do not necessarily use them. On the other hand, in Alma Swan's 2004 study<sup>5</sup>, 69% of the researchers surveyed said that if self-archiving ever became compulsory they would comply willingly, and a further 8% unwillingly; in 2005<sup>6</sup> this figure had risen to 81% and 13%. A far more significant factor, however, is the growing number of research funders – and indeed universities and other institutions – that are now leaning towards self-archiving policies of one kind or another. Some of these are compulsory, others are voluntary; but they are growing in number and that trend is probably the most significant influence on what is happening.

What is the likely effect on journals? The sense in which I say that institutional repositories are parasitical on journals is that they take advantage of the journal system; they point to published articles even when they do not reproduce the final

published versions. (Of course Paracite<sup>7</sup>, the system which was set up to trace citations in repositories, was not named by mistake; whoever named it certainly thought they were parasitic and I think that's significant.)

## The likely effect on journals

### Cancellations

Two recent surveys of librarians' views, by Ware<sup>8</sup>, and Beckett and Inger<sup>9</sup>, both indicate very clearly that when a sufficient percentage of the final version of author articles is easily and freely available in repositories, cancellations will follow. Table 1 compares some of the key elements of those reports.

In Mark Ware's survey 81% of respondents said that availability in a repository would be 'very important' or 'important' in making cancellation decisions; however, they said that pricing and usage were far more important factors in cancellation, followed by user needs. They did not see either

Ware (March 2006)	Beckett and Inger (October 2006)
340 responses	424 responses
81% said availability in an OA repository would be a 'very important' or 'important' factor in cancellation decisions (but behind pricing [95%], usage [95%], user needs [93%])	'a significant number of librarians are likely to substitute OA materials for subscribed resources, given certain levels of reliability, peer review and currency'
Preprint/postprint versions not seen as adequate substitute (but PDF is)	Author's unrefereed, uncorrected original MS is least adequate substitute Post-peer review version (irrespective of publishers' editing) is adequate
32% think publishers should not be worried	8% think publishers should not be worried
11% think publishers should be worried	38% think publishers should be worried
54% think it is too early to tell	

Table 1. Comparison of reports by Ware, and Beckett and Inger on librarians' views on repositories

preprints or postprints as being adequate substitutes for the published version, though, unsurprisingly, a PDF copy of the final version was considered adequate. 32% thought that publishers should not be worried; 11% thought that they should be worried; and 54% thought it was just too early to say.

The Beckett and Inger study was carried out only a few months later; they concluded that a significant number of librarians were likely to substitute open access materials for subscribed resources given certain levels of reliability, peer review and currency. The respondents found the authors' preprints – the uncorrected original scripts – the least adequate as a substitute for the published version; the version after peer review, whether or not it had incorporated the publisher's editing, was adequate. A similar number to those in the Ware study thought that publishers should not be worried, but a very much larger number thought that they should be worried; I think publishers should be worried too! Unfortunately we will not know for sure whether repositories kill journals unless and until they do.

It is too early to be able to study the effect on subscriptions where journals have their full content in repositories. However, what we can do is look at the example of a number of journals that have made their own content free, although of course this is not an exact comparison with self-archiving. The *British Medical Journal*<sup>10</sup> used to make all its content free on publication; the publisher found that both print subscriptions and advertising revenue fell dramatically. As a result, the policy was changed; research articles are still free but none of the other content is free and, as a result, revenue has almost recovered to its previous level. The publisher of *Molecular Biology of the Cell*<sup>11</sup> decided to make all the journal's content freely available

after two months; in the following three years its annual subscription growth fell catastrophically from a spectacular 84% to just 8%. The publisher of *Proceedings of the National Academy of Science*<sup>12</sup> introduced a policy in 2000 where the content was free after a month; and in the following year subscriptions fell by 11%. That period was then extended from one month to six months; this resulted in a slight recovery, to 9% decline rather than 11%.

None of us has a huge amount of data yet, one way or another, but these few examples do seem to indicate that there may be serious dangers. I have to say that it seems logical to me: if a librarian is confronted with a need to cancel some of the journals you take, won't she or he cancel the ones where all or almost all of the content is free somewhere, rather than the ones where it's not? We know that librarians are having to make cancellations because they do not have enough money to buy everything they would like; one of the factors to take into account, surely, is whether (all or most of) the content is freely available somewhere? This is actually borne out by a recent study from the Research Information Network<sup>13</sup>. The authors say that 'all funders are aware that deposits of significant numbers of articles in repositories could lead to loss of subscriptions to established journals and the danger that this may pose to learned societies in particular.' So it is not just publishers who think this.

#### Usage

The London Mathematical Society, some of whose journals have a large number of articles deposited in arXiv<sup>14</sup>, noted that it has 23% fewer downloads on its own site for those articles that were also available in arXiv;<sup>15</sup> it seems to me that a 23% drop in downloads is quite serious. Remember that 95%

of librarians said that usage, along with pricing, was the most important factor in determining cancellations. The Institute of Physics also says<sup>16</sup> that for those journals whose content is largely mirrored in arXiv, it sees a very significant drop in usage on its own journals website.

It is also interesting to wonder why, when they have arXiv and they also have extremely good journal access, the high energy physics community is making such an active move towards open access publication, with the sponsorship consortium that they are setting up<sup>17</sup> to make sure that they can convert existing journals to open access where authors do not have to pay. One reason for doing that might be a concern that otherwise the journals will suffer because arXiv so completely covers the content of some of them.

### Humpty Dumpty

Leo Walford recently wrote a very interesting article<sup>18</sup>, making the analogy that Humpty Dumpty looked fine on the way down, but after he had hit the ground, nobody could put him together again. While we cannot be sure that there will be damage to journals as we know them (and to their publishers), there are pointers suggesting that this might be the case; and by the time the damage has occurred it will be too late to undo it. Thus it is essential that we are conscious now of the possible consequences of the move towards self-archiving, and that we consider what we are going to do about it.

### Does it matter?

If a lot of subscription journals are damaged by the fact that their content is all or mostly self-archived, will it actually matter? Let us consider some of the possible consequences.

#### Peer review

Of course journals do not do peer review – academics do peer review. However, journals currently provide the framework for carrying out peer review, and academics value this extremely highly. In an ALPSP study of authors<sup>19</sup> in 2002 (and I'm sure it's still true), 96% 'strongly agreed' or 'agreed' that they preferred to submit articles to journals that maintained peer review. Throughout

the academic community peer review is considered to be hugely important, although its flaws are recognized and people are looking at other ways of doing it. The *Nature* experiment with open peer review<sup>20</sup> was interesting; only 5% of authors took up the option and it has now been abandoned, but *Nature* is to be commended for experimenting. There have also been proposals that in future peer review might be applied directly to the content of repositories; I think the first person to advocate this was probably John Smith in 1998,<sup>21</sup> so these ideas have been around for a while but it has not really happened yet.

#### Editing

One of the functions of journals is copy-editing to improve the clarity and readability of journal articles; in addition (and this is a factor which has become increasingly important), copy-editors check that the references are accurate, complete and properly formatted to enable linking from the references to the citations. Surveys<sup>22</sup> have shown that reference linking is one of the most valuable new features that electronic versions of new journals bring; but if the references are not correct, the linking does not work. I have found since I've become a journal editor that getting the references right takes up a lot of my time because the number of authors (even when they are publishers) who actually get all their references right is surprisingly low!

A recent study by Edward Wates and Robert Campbell at Blackwell<sup>23</sup> compared authors' and the publisher's versions of articles; they found that 42.7% of queries from editors were to do with the references being either wrong or incomplete, 13.6% were about data that was actually missing, and 5.5% led to alterations that really altered the sense of what the article said. A more recent article by David Goodman and his colleagues at Long Island University<sup>24</sup> has looked at 12 articles in biochemistry and 12 in social science to identify the differences between the self-archived version and the published version. In the biochemistry articles they found no differences which affected the validity of the findings; in the social sciences, two. It is, of course, a very small sample, but the findings are fairly similar to those of Wates and Campbell, who looked at 189 articles. Goodman *et al* did find, however, that readability was greatly improved in the publishers' versions. One thing that concerns me is that if we move to a system where we do not

have journals – or something like them – to carry out the editing function, then those authors whose writing is most in need of editing, who will be those who are not writing in their own native language and who may therefore be least able to pay for a publication charge for the publication services of an open access journal, may be seriously disadvantaged. It may well be that Western authors are intelligible more or less – again, my experience as an editor suggests sometimes it is less rather than more – but authors who are writing in a completely different language from their own may be very hard to understand and the value of their work may actually get lost.

One of the problems with editing, though, is that it is invisible. If you have done a good job as an editor nobody knows you have done it; they simply think that the authors are better than they are. In 2002 ALPSP conducted a survey of academics<sup>25</sup>, asking them to respond both as authors and as readers. One of the questions was to do with what characteristics of journals as we know them needed to be preserved in any new system. In their author role, 60% of respondents thought that content editing and improvement of articles should be maintained, 50% that language copy-editing should be maintained, and 46% checking citations and adding links to citations. As readers, on the other hand, the figures were lower: 39% for content editing, 34% for language editing and 28% for citation checking. This demonstrates that authors, the people who are aware of the editing that happens to their work, do value it quite highly. Worryingly, however, not all journals do as much copy-editing as they used to; some outsource it to countries where people may themselves not be native English speakers (some of

whom do a very good job) but others are cutting right back on copy-editing in the interests of cost-saving.

#### *The journal package*

Another thing that journals do is to provide a convenient package. They select and collect together content that is particularly relevant and interesting to a specific community or sub-community. The more information there is, and the more overloaded an area is, the more important it is that somebody collects together the content that the time-pressed researchers actually want to read, although it would be a mistake to assume that journals are the only way to do this.

#### *What happens to the money?*

Another feature of the journal system as we know it is what happens to the proceeds from journal publishing at present. It has been pointed out, by Raym Crow amongst others,<sup>26, 27</sup> that by no means all publishers and all journals are commercial. As Figure 1 shows, 45% of journals are both owned and published by commercial publishers. 38% are non-profit journals published by their own organizations. The remaining 17% are non-profit journals which are published under contract by commercial publishers. Thus 55% of journals belong to non-profit organizations.

As we know, non-profit organizations put their money back into other activities to the benefit of their communities. In 2004 ALPSP and Blackwell surveyed a number of learned societies.<sup>28</sup> 33% of them put some of their publishing surpluses towards keeping the cost of conferences down; 32% to keeping membership fees as low as possible; 26% put money into public education;

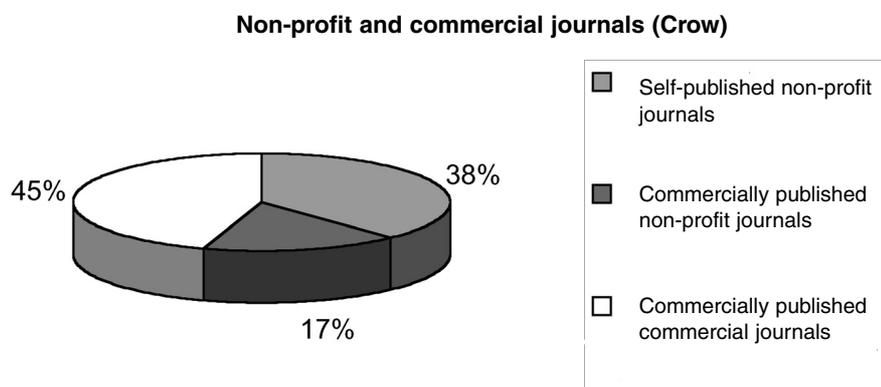


Figure 1. Non-profit and commercial journals – Ulrich's Periodicals Directory, 2005 (analysis by Raym Crow)  
Reproduced with permission from 'Learned Publishing'

26% provided bursaries for people to attend their own and other people's meetings; and 21% put money into their own research funding. All those areas are thus dependent upon there being a flow of money from journal publishing. But many of the arguments about changing the journal publishing model tend to assume that journal profits (or surpluses) will reduce as a consequence. However, if that happens, these benefits will reduce too, and that will have a knock-on effect on the scholarly community. Maybe that is what we want, but let's think about it.

Some journals may be more vulnerable than others in this situation; publishers with only one or a few journals may be particularly vulnerable. Crow<sup>29</sup> says that over 97% of society publishers publish three or fewer, and almost 90% have just one journal, so society publishers are especially vulnerable. They are also, of course, restricted to the discipline of their society; the Royal Society of Chemistry cannot suddenly decide to publish a mathematics journal because that might help to keep it afloat! They therefore tend to have limited room for manoeuvre. In addition, journals whose profit margin is already thin are at greater risk; and it is generally believed<sup>30</sup> (although I am not aware of any published comparisons) that non-profit publishers tend to make less money than commercial publishers – certainly the average society surpluses found by Baldwin<sup>31</sup> were quite low. They thus also have less of a financial cushion to deploy if they want to change their plans.

Niche journals, which may or may not come from societies, are also vulnerable because with their low circulation they are bound to have a higher price;<sup>32</sup> if a journal is only going to sell four or five hundred copies the publisher will obviously have to charge a great deal more than for a journal which will sell four or five thousand. And as Ware<sup>33</sup> found, price is one of the key factors when a librarian decides to cancel. Thus niche journals are particularly vulnerable.

### What should publishers do about it?

To start with, what should publishers not do about it? Simply arguing that repositories are a bad thing is not going to do any good at all; in fact it could do more harm than good to the industry – it just sounds like 'shroud-waving'. For publishers to appear to be hostile to what is happening cannot

do any good to the industry or its image. Some of you may remember the book *1066 And All That*<sup>34</sup>, where the Cavaliers were 'Wrong but Wromantic' and the Roundheads were 'Right and Repulsive'. It would be a very bad thing if publishers ended up being Right and Repulsive!

So what should we do about it? First of all, we need to make sure that the communities with which we engage are clear about the possible – and likely – consequences of widespread, particularly mandatory, self-archiving. They need to understand that the solutions to those will not be the same everywhere. One size does not fit all – subjects are different; journals are different. Journals have different frequencies, and have different constituencies. Some journals have the same community of writers as of readers in very research-heavy disciplines, while others, for example in clinical medicine, have a fairly small community of research authors and a very large community of readers who would never do any research themselves. But the information we need if we are to make people aware of the likely consequences has to be based on factual evidence; so research has to continue into what actually happens as self-archiving mandates increase and begin to bite.

In addition, publishers need to make their content as available as they can without destroying their business. Again, one size does not fit all; it will not be the same for every publisher or for every journal. Publishers might wish to look at whether or not they can switch to some form of open access model; David Prosser<sup>35</sup> has argued that the hybrid, or 'author choice', model is a possible way for a publisher to dip a toe into the water to see what the response is from authors. It is, of course, absolutely crucial for publishers to get their sums right: charge too much and no authors will adopt the model; charge too little and, if they all adopt it, the publisher will go out of business. If a publisher does not feel it can switch to an open access model, it will need to decide whether it is necessary to impose an embargo period before people can self-archive in order to protect subscriptions and if so, how long should it be: one month, six months, one year? It is hard to determine the correct period. All a publisher can do is to look at the available data.

More than this, though, publishers need to think about new ways of adding value to scholarly communication. I think, actually, that in worrying about open access and journals we may be looking

in the wrong place; we may be worrying about the spider in one corner of the room when there's a tiger in the other corner – it's about time we turned round and looked at the tiger!

A really dramatic, and potentially far more challenging, change is taking place: scholarly communication is turning into something very different (and much more exciting). But although journals as we know them may have a place in the new structure, it is only a fairly small place. The way that researchers are creating, using and interacting with data, and the way that they are communicating with each other in a whole range of ways from the ultra-formal (as in journals) to the ultra-informal (as in blogs and wikis), is a far more important revolution than anything that may happen to the business model of the journal on its own. Publishers have to think about adding value to scholarly communications in completely new ways if they are going to stay in business at all in the longer term – they have to 'ride the wave'.

Resistance is futile; publishers cannot say that self-archiving and repositories shouldn't happen because they will hurt them – they are happening. Publishers cannot stop this development; they must go with it and see how they can make a creative business that actually helps scholarship in the future. It is vital, therefore, to understand what journals are for and who they serve – both the authors and readers whom they serve directly, and also the funders and universities and other institutions whose interests they serve in a variety of different ways. Publishers and all those who benefit from journals in these different ways need to collaborate to figure out what are the functions that journals do currently carry out, which of these are really important, which are important enough to keep and what is the best way of doing it. We should not forget the fate of the man with the red flag who used to walk in front of cars when they were first invented – like him, publishers might not have a job if they insist on being publishers as we know them today.

## Conclusions

Institutional repositories are not going to go away. They do have the potential, at least, to do great damage to many journals, if not all, and it is therefore essential to make funders and others aware of what evidence there is. However, publishers must avoid

shroud-waving and saying 'this is terrible, you're going to kill us, don't do it'. That is simply not going to work. Instead, publishers need to work with the communities that they serve, to figure out how best to add the value that they really do want. Publishers should not assume that there is a role in future either for journals as we know them or for publishers as we know them.

## References

1. Davis, P and Connolly, M J L, Institutional Repositories: Evaluating the Reasons for Non-use of Cornell University's Installation of DSpace, *D-Lib Magazine*, 2007, 13(3/4). Available at <http://www.dlib.org/dlib/march07/davis/03davis.html>
2. Kennan, M A, Academic authors, scholarly publishing, and open access in Australia, *Learned Publishing*, 2007, 20(2), 138–146. <http://dx.doi.org/10.1087/174148507X185117>
3. Wigand, R T and Hess, T, Open Access & Science Publishing: Results of a Study on Researchers' Acceptance and Use of Open Access Publishing. Report by Ludwig-Maximilians-Universität Munich, Germany, Institute for Information Systems and New Media and University of Arkansas at Little Rock, USA, Department of Information Science. Available at [http://openaccessstudy.com/Hess\\_Wigand\\_Mann\\_Walter\\_2007\\_Open\\_Access\\_Management\\_Report.pdf](http://openaccessstudy.com/Hess_Wigand_Mann_Walter_2007_Open_Access_Management_Report.pdf)
4. Van Westrienen, G and Lynch, C A, Academic Institutional Repositories: Deployment Status in 13 Nations as of Mid 2005. *D-Lib Magazine*, 2005, 11(9). Available at <http://www.dlib.org//dlib/september05/westrienen/09westrienen.html>
5. Swan, A and Brown, S (2004). JISC/OSI Journal Authors Survey Report. In *JISC Report*. Available at [http://www.jisc.ac.uk/uploaded\\_documents/JISCOAreport1.pdf](http://www.jisc.ac.uk/uploaded_documents/JISCOAreport1.pdf)
6. Swan, A and Brown, S (2005). Open access self-archiving: an author study. Technical Report, Joint Information Systems Committee (JISC), UK FE and HE funding councils. Available at <http://cogprints.org/4385/01/jisc2.pdf>
7. <http://paracite.eprints.org/>
8. Ware, M, Factors in Journal Cancellation. Worthing, Association of Learned and Professional Society Publishers, 2006.

9. Beckett, C and Inger, S, Self-archiving and Journal Subscriptions – co-existence or competition? London, Publishing Research Consortium, 2006. Available at [http://www.publishingresearch.net/self\\_archiving.htm](http://www.publishingresearch.net/self_archiving.htm)
10. Personal communication
11. Everngam, R, *Delayed Free Access: the experience of Molecular Biology of the Cell*. Presentation at ALPSP/SSP seminar, 'Open Access Publishing: Does it really work in practice?', 8 November 2004. Available at [http://www.sspnet.org/documents/52\\_Everngam.pdf](http://www.sspnet.org/documents/52_Everngam.pdf)
12. Personal communication, quoted with permission at <http://www.library.yale.edu/~llicense/ListArchives/0612/msg00033.html>
13. Research Funders' Policies for the Management of Information Outputs. London, Research Information Network, 2007. Available at <http://www.rin.ac.uk/policy-information-outputs>
14. <http://www.arxiv.org/>
15. Davis, P M and Fromerth, M J, Does the arXiv lead to higher citations and reduced publisher downloads for mathematics articles? *Scientometrics*, 2007, 71(2). <http://dx.doi.org/10.1007/s11192-007-1661-8>
16. See <http://users.ecs.soton.ac.uk/~harnad/Hypermail/Amsci/4836.html>
17. See <http://press.web.cern.ch/press/PressReleases/Releases2006/PR16.06E.html>
18. Walford, L, Personal View: Humpty Dumpty. *Learned Publishing*, 2007, 20(2), 154–155. <http://dx.doi.org/10.1087/174148507X185135>
19. Swan, A, *Authors and Electronic Publishing*. Worthing, Association of Learned and Professional Society Publishers, 2002.
20. See <http://www.nature.com/nature/peerreview/debate/nature05535.html>
21. Smith, J W T, The deconstructed journal – a new model for academic publishing. *Learned Publishing*, 1999, 12(2) 79–91. Available at <http://www.ingentaconnect.com/content/alpsp/lp>
22. Pullinger, D and Baldwin, C, *Electronic Journals and User Behaviour: learning for the future from the SuperJournal Project*. Cambridge, deedot press, 2002.
23. Wates, E and Campbell, R, Author's version vs. publisher's version: an analysis of the copy-editing function. *Learned Publishing* 2007, 20(2), 121–129. <http://dx.doi.org/10.1087/174148507X185090>
24. Goodman, D, Dowson, S and Yaremchuk, J (2007). Open access and accuracy: author-archived manuscripts vs. published articles, *Learned Publishing*, 2007, 20, 203–215. <http://dx.doi.org/10.1087/095315107X204012>
25. Swan, A, *Authors and Electronic Publishing*. Worthing, Association of Learned and Professional Society Publishers, 2002.
26. Crow, R, Publishing Cooperatives: an alternative for society publishers (2006). *First Monday*, 2006, 11(9). Available at [http://www.firstmonday.org/issues/issue11\\_9/crow/index.html](http://www.firstmonday.org/issues/issue11_9/crow/index.html)
27. Morris, S, Mapping the journal publishing landscape: how much do we know? *Learned Publishing*, 2007, 20, 299–310. <http://dx.doi.org/10.1087/095315107X239654>
28. Baldwin, C, What do Learned Societies do with their Publishing Surpluses? Worthing, Association of Learned and Professional Society Publishers, 2004. Available at [http://www.alpsp.org/ngen\\_public/article.asp?id=200&did=47&aid=277&st=&oaid=-1](http://www.alpsp.org/ngen_public/article.asp?id=200&did=47&aid=277&st=&oaid=-1)
29. Crow, R, Publishing Cooperatives: an alternative for society publishers (2006). *First Monday*, 2006, 11(9). Available at [http://www.firstmonday.org/issues/issue11\\_9/crow/index.html](http://www.firstmonday.org/issues/issue11_9/crow/index.html)
30. Bergstrom, C T and Bergstrom, T C, The economics of scholarly journal publishing (2001, updated 2006). Available at <http://octavia.zoology.washington.edu/publishing/>
31. Baldwin, C, What do Learned Societies do with their Publishing Surpluses? Worthing, Association of Learned and Professional Society Publishers, 2004. Available at [http://www.alpsp.org/ngen\\_public/article.asp?id=200&did=47&aid=277&st=&oaid=-1](http://www.alpsp.org/ngen_public/article.asp?id=200&did=47&aid=277&st=&oaid=-1)
32. King, D W, The cost of journal publishing: a literature review and commentary, *Learned Publishing*, 2007, 20(2), 85–106. <http://dx.doi.org/10.1087/174148507X183551>
33. Ware, M, *Factors in Journal Cancellation*. Worthing, Association of Learned and Professional Society Publishers, 2006.
34. Sellar, W C and Yeatman, R J, *1066 And All That*. London, Methuen, 1930.

35. Prosser, D C, Between a rock and a hard place: the big squeeze for small publishers. *Learned Publishing*, 2004, 17(1), pp. 17–22.  
<http://dx.doi.org/10.1087/095315104322710214>

Article © Sally Morris

---

■ **Sally Morris**  
Consultant, **Morris Associates (Publishing Consultancy)**  
South House, The Street  
Clapham, Worthing, West Sussex BN13 3UU, UK  
Tel: +44(0)1903 871286  
Fax: +44(0)8701 202806  
E-mail: [sally@morris-assocs.demon.co.uk](mailto:sally@morris-assocs.demon.co.uk)

---

To view the original copy of this article, published in *Serials*, the journal of the UKSG, click here:

<http://serials.uksg.org/openurl.asp?genre=article&issn=0953-0460&volume=20&issue=3&page=172>

For a link to the table of contents for the issue of *Serials* in which this article first appeared, click here:

<http://serials.uksg.org/openurl.asp?genre=issue&issn=0953-0460&volume=20&issue=3>