
Access from the University of Nottingham repository:
http://eprints.nottingham.ac.uk/706/1/Serials_2007_edited_version.pdf

Copyright and reuse:
The Nottingham ePrints service makes this work by researchers of the University of Nottingham available open access under the following conditions.

This article is made available under the University of Nottingham End User licence and may be reused according to the conditions of the licence. For more details see:
http://eprints.nottingham.ac.uk/end_user_agreement.pdf

A note on versions:
The version presented here may differ from the published version or from the version of record. If you wish to cite this item you are advised to consult the publisher’s version. Please see the repository url above for details on accessing the published version and note that access may require a subscription.

For more information, please contact eprints@nottingham.ac.uk
Can open access repositories and peer-reviewed journals coexist?

Published in Serials 20 (3) 2007, 163-171

Based on a paper presented at the 30th UKSG Conference, Warwick, April 2007

STEPHEN PINFIELD
Chief Information Officer, University of Nottingham

Abstract
It is often assumed that open access repositories and peer-reviewed journals are in competition with each other and therefore will in the long term be unable to coexist. This paper takes a critical look at that assumption. It draws on the available evidence of actual practice which indicates that coexistence is possible at least in the medium term. It discusses possible future models of publication and dissemination which include open access, repositories, peer review and journals. The paper suggests that repositories and journals may coexist in the long term but that both may have to undergo significant changes. Important areas where changes need to occur include: widespread deployment of repository infrastructure, development of version identification standards, development of value-added features, new business models, new approaches to quality control and adoption of digital preservation as a repository function.

Introduction
It is often assumed that open access (OA) repositories and peer-reviewed journals are in competition. In this competitive situation, it is suggested, coexistence in the long term will be impossible and only one of the two will exist in the scholarly communication environment. However, this paper takes a critical look at that assumption by addressing the question:

Are repositories and journals necessarily in competition?

In order to answer this, the paper also addresses a number of sub-questions:

Is there any current empirical evidence of either competition or coexistence?
What are the implications of this evidence for the short, medium and long term?
What possible future publishing and dissemination models are there for repositories and journals which might involve coexistence?

This paper adopts the working definition of an open access repository as a set of systems and services which facilitates the ingest, storage and management of digital objects, and where those objects can than be freely and immediately accessed and reused in an unrestricted way. Repositories may be maintained by and serve different communities, such as institutions or subject groupings. They may also hold different types of content, including copies of research papers, digital learning objects, theses and data. This paper will concentrate primarily on the first of these, copies of research papers, sometimes referred to as ‘e-prints’.
E-prints are electronic copies of journal articles and similar research outputs. They may take the form of ‘preprints’ (versions of papers before they have been refereed) or ‘postprints’ (versions of papers which include changes made in response to referees’ comments, normally in the format produced by the author).

**Empirical evidence**

Perhaps some of the most important empirical evidence relating to the question of the relationship between repositories and journals is provided by services that already exist and have been operating for some time. The arXiv repository is a prominent example of this. As it has been in existence since 1991, it provides ‘real life’ evidence for a particular set of disciplines (physics, mathematics, computer science and quantitative biology). It now holds over 415,000 papers and in key areas (high energy physics, astrophysics, and condensed matter) includes copies of a large proportion of the available journal literature, which is in both the form of ‘preprints’ and ‘postprints’.

So how has arXiv affected peer-reviewed journals? Two major areas need to be examined: usage and subscriptions.

Recent data on usage has been produced by Edwin H Henneken et al. In the area of Astrophysics they analysed papers first made available as e-prints in August 2004 and then published in four core peer-reviewed journals. They looked at ‘reads’ per paper for the period between August 2004 and June 2006 from arXiv and the NASA Astrophysics Data System (ADS) on the one hand and the journals on the other. Their data shows clearly that the usage of material in repositories is concentrated in the period before formal publication of the journal article. Once an article is formally published, most usage switches to the journal site. The ‘half life’ of an e-print (the period of time it takes for usage to fall to a half of its highest point) is very short. ‘The typical users prefer to read the journal article when this becomes available’, they state. This is a situation of coexistence, rather than competition. They conclude:

‘This is good news for the publishers. Eprints have not undermined journal use in the astrophysics community and thus do not pose a threat to the journal readership.’

Published data on subscriptions is a little harder to come by, but what there is in the public domain also points towards coexistence rather than competition. In a study on the open access repositories, Alma Swan states:

‘...we asked the American Physical Society (APS) and the Institute of Physics Publishing Ltd (IOPP) what their experiences have been over the 14 years that arXiv has been in existence. We asked how many subscriptions have been lost as a result of arXiv. Both societies said they could not identify any losses of subscriptions for this reason. Subscription movements for the journals they publish in the areas covered by arXiv are no different from those of their journals in other areas of physics over the period. Moreover, both societies say that they do not view arXiv as a threat to their business (rather the opposite, in fact) and this is underlined by the fact that the APS helped establish an arXiv mirror site at the Brookhaven National Laboratory – hardly the action of a society with its back to the wall because of that repository. Now it is true that there are only a couple of experiments of this sort carried out so far (physics and computer science), where publishers have to co-exist with a successful open access archive, and so there is always the possibility that there is
something of a ‘special case’ about this example. Quite what might make it such a special case has never been adequately argued, but it is a finite possibility. Nevertheless, the evidence there is to hand points to the likelihood that the peaceful — and perhaps mutually beneficial — co-existence of traditional journals and open access archives is entirely possible; in biological terms, mutualism, rather than parasitism or symbiosis, might best describe the relationship.’

Data presented by Andrew Wray\(^3\) of the Institute of Physics (IOP) to the OAI 5 conference seems to concur with Swan’s conclusions. Wray shows that data gathered by the IOP indicates that arXiv is not having any major impact on IOP journal subscriptions. He also outlines ways in which the IOP is working with arXiv to create new innovative services (including www.eprintsweb.org) for physicists.

The evidence that subscription patterns of relevant journals have not been significantly affected during the lifetime of arXiv reinforces the evidence that usage patterns of journals and arXiv seem to be complementary. The data points to coexistence rather than competition, at least for the medium term.

Swan’s question of whether or not arXiv is a ‘special case’ does, however, require consideration. Because of the significance of arXiv, various attempts have been made to try to define how high energy physics in particular is a unique discipline and therefore how arXiv might be a special case. It has been suggested, for example, that physicists who use arXiv are a small community of researchers who tend to work in teams and cluster around a small number of large pieces of scientific equipment (such as the facilities at CERN). The argument is put that this community values peer-reviewed journals less than other subject communities since formal peer review itself is less important to its members. Most researchers, it is suggested, know each other personally (or at least by reputation), and therefore have less need for the filtering and quality assurance that peer review provides. These factors, it is argued, make arXiv a special case. However, these arguments are questionable on a number of levels. To begin with, it is clear that arXiv is used by more people than just high energy physicists carrying out experiments on large-scale equipment. Many theoretical physicists and astronomers in a wide range of institutions use it. This fact tends to undermine the argument that arXiv is only used by a small tightly-knit community.

Furthermore, data produced by Henneken et al demonstrates that arXiv users are interested in peer review. They make extensive use of peer-reviewed journal articles in preference to e-prints when the articles become available. This point is further illustrated by the fact that it is this same community that is now making a concerted effort to transition its key peer-reviewed journals into open access publications. This initiative, led by major stakeholders such as CERN, is being pursued even though the community makes extensive use of arXiv. The community clearly values peer-reviewed journals as well as OA repositories.

The arXiv repository is not unique. Successful repositories exist in a small number of subject disciplines, particularly physics, computer science and economics. So, are these all special cases? It has been pointed out that all these disciplines had pre-existing preprint cultures, a fact which it is suggested has made them more inclined than other disciplines to adopt e-print repositories. Other disciplines without preprint cultures, it is argued, have shown far less inclination to adopt repositories. However, adopting e-print repositories does not necessarily mean adopting a preprint culture. Repositories do not necessarily have to hold preprints. Many disciplines may prefer repositories they use to only hold papers which have already been through a recognized quality-control process.
It is not clear how a ‘special case’ argument can be sustained for computer science and economics as well as physics. Neither is it clear how the growing range of institutional repositories (IRs), most of which are cross disciplinary, fit with the ‘special case’ argument. Some IRs, for example, in The Netherlands, are just beginning to show signs of critical mass. It is getting difficult to maintain a meaningful special-case argument for arXiv in the context of these developments.

However, other lines of evidence are more equivocal. Survey data on the attitudes of key stakeholders in the scholarly communication chain is an example of this. Mark Ware, Chris Beckett and Simon Inger have provided evidence on the attitudes of librarians both of which point to considerable uncertainty in the profession about these issues and their future impact. Ware reports that although OA repositories are not currently considered to be a substitute for journals, 53% of respondents (rising to 81% in five years) indicate that content in repositories would be an important (or very important) factor in determining journal cancellations. Beckett and Inger report a number of key conclusions:

‘As many as 40% [of respondents] believe that libraries are wasting their money subscribing to journals when almost the same content is available for free on repositories; but a similar proportion disagree.

There is concern about the impact repositories will have on journals viability, though 31% believe it will have no impact.

Just a third agree that Open Access will impact negatively on low-quality journals only, implying that it will also impact negatively on high-quality journals.

A minority (just 26%) believe that journals will be forced to charge authors and more believe this won’t be the case.’

However, this data leads to less definite conclusions. What it shows in fact is that opinions are divided amongst practitioners. There is considerable uncertainty about what the long-term picture will look like. Beckett and Inger acknowledge this in relation to some of their conclusions:

‘Opinions are clearly split over whether libraries are wasting their money subscribing to journals when almost the same content is available for free on repositories (40/41). This result fails to provide any reassurance to publishers that there is no threat of substitution of content in an OA archive for the journal subscription.’

This last inference is certainly the case. But it is also true that neither does the result justify any short-term panic amongst publishers.

This data on attitudes is important but needs to be treated with some caution. The surveys quoted look at current attitudes and ask respondents to anticipate their future attitudes and behaviours. It is, of course, difficult for people to predict reliably what their attitudes will be in the future. It can be even more difficult for them to predict how their possible future attitudes will affect their behaviour, especially when there is usually a gap between expressed attitudes and actual behaviour (what people say against what people do) even in the present. The data cannot therefore be treated as an entirely reliable view of the future.

In summary, empirical evidence of actual behaviour both in terms of usage and subscriptions provides evidence of coexistence rather than competition in the
short and medium term. Other evidence of attitudes, however, presents a more equivocal picture, indicating at the very least that there is uncertainty about how the scholarly communication environment will develop. It is essential, therefore, that empirical evidence of all sorts continues to be gathered and made public so that it can be used to inform the ongoing discussion. For the present, however, it can be concluded that the available data seems to indicate that OA repositories and peer-reviewed journals are not necessarily in competition.

Future scenarios

Whilst empirical evidence seems to point towards the coexistence of open access repositories and peer-reviewed journals in the short and medium term, the question of the long-term future remains. Drawing any definite conclusions on this question is, of course, impossible. Any industry closely connected with information and communication technology will inevitably undergo a number of profound changes over the next decade and the scholarly communication industry is no exception. Different kinds of content, probably making use of rich media, will become increasingly important in research communication and this is likely to lead to profound changes in communication norms. However, mapping possible future models, with reference to the roles of journals and repositories, may provide a useful framework within which discussion of the future of research communication can move forwards.

Despite technological change, it is likely that the basic functions of scholarly communication will remain relatively constant. These functions are normally expressed as:

- registration
- certification
- dissemination
- archiving.

Registration is associated with defining and recording responsibility for a piece of work in a public way. This may often be connected with asserting priority – ensuring that work (and the ideas behind it) can be correctly attributed to a particular person or group. Certification is the quality assurance process which marks certain works as having been through particular quality-control processes, normally peer review. Dissemination is about the circulation of the work so that it finds its readership and makes an impact. Archiving relates to preserving long-term access to the content so that there is a reliable record of scientific and scholarly findings which can be read, cited and built on in the future.

At present, all these functions are most commonly carried out by different parties working with peer-reviewed journals, although in certain disciplines conference proceedings, essays in edited books and monographs remain important.

Bearing in mind these functions, there are perhaps four possible future models or scenarios of scholarly communication:

1. Journals remain the primary means of scholarly communication and repositories are not significant.
2. Journals and repositories coexist – with no changes to current business models.
4. Repositories displace journals as the primary means of scholarly communication.
Two of these scenarios (2 and 3) involve the coexistence of journals and repositories in the scholarly communication arena. The other two models involve the dominance of either one or other.

Scenario 1 in most disciplines is the status quo. It does not necessarily mean the repositories will have no significance at all, simply that their significance will not lie in delivering scholarly content currently published in journals. It is possible that repositories will find a complementary function delivering content such as data, grey literature and dissertations. However, in this model, the main scholarly communication functions of registration, certification, dissemination and archiving remain with the journal.

Scenario 2 implies that repositories will become more important than at present but that they will not undermine the position of journals nor impact upon current subscription-based business models. In this scenario, journals and repositories both perform the registration function, although it is often thought that repositories have the potential to become most important in this area since they allow material to be put into the public domain very quickly. Journals continue to perform certification. Dissemination and archiving are performed by both but journals retain the dominant part in the latter. Something like this view is held by Stevan Harnad who normally seems to be reluctant to discuss the possibility that the growth of repositories may lead to changing business models associated with scholarly content delivery.

Scenario 3 also involves the coexistence of journals and repositories but involves journals adopting new business models. The most obvious business model is the publication-charge model already used by many OA journals, such as Public Library of Science titles, in which an input-side charge is made by the publisher when an article is submitted or accepted in order to pay for the costs of publication. The article is made freely available when published and can then be deposited in a repository. A hybrid business model, where publishers give authors an OA option for a particular paper published in a subscription journal on payment of a per-article fee, might also fit with this scenario. This hybrid model is favoured by Robert Terry of the Wellcome Trust as part of his promotion of the Wellcome policy to put all Wellcome-funded papers in PubMed Central (PMC). In this scenario, certification continues to be provided by journals but the other scholarly communication functions (registration, dissemination and archiving) are shared between journals and repositories. Supporters of this approach would argue that dissemination and archiving can increasingly be performed by a well-funded central repository such as PMC.

Scenario 4 is the most radical. Like Scenario 1 it does not involve the coexistence of journals and repositories in the scholarly communication field. Unlike Scenario 1, in this scenario, the key scholarly communication functions of registration, certification, dissemination and archiving are removed from journals entirely and are performed in other ways associated with repositories. The possibilities of such a scenario have been discussed by Raym Crow and Jean-Claude Guédon but go back to earlier discussions by John Smith and others. The most profound change here is that quality control is carried out outside the remit of journals. In principle, of course, there is no reason why this should not happen, but in practice significant cultural and operational barriers would need to be overcome before such a scenario could become widespread.
Possible changes
Scenarios 2 to 4 all require change to the current system of scholarly communication if they are to be workable. Key changes include the following:

- widespread deployment of repository infrastructure
- development of version identification standards
- development of value-added features
- new business models
- new approaches to quality-control
- adoption of digital preservation as a repository function.

These different issues are discussed below.

Widespread deployment of repository infrastructure
Over the last two to three years the number of OA repositories has grown rapidly. In June 2007, OpenDOAR listed more than 900 OA interoperable repositories (that is, populated OA repositories that are compliant with the Open Archives Initiative Protocol for Metadata Harvesting). Such repositories are being set up and maintained by research institutions, national libraries, subject communities and research funders. However, a great deal remains to be done.

Firstly, a large number of other repositories need to be set up to provide anything like coverage for the global research community. In the UK, organizations such as the Joint Information Systems Committee (JISC) are encouraging and funding the setting up of repositories in institutions through a succession of development programmes. This includes the Repositories Support Project, led by the University of Nottingham, which is providing proactive support for higher education institutions (HEIs) by setting up repositories. The JISC has also funded The Depot, a national ‘interim’ repository. The Depot has been designed to house material produced by researchers in UK HEIs where they do not yet have access to a repository maintained by their institution. It is anticipated that The Depot will transfer material it holds to institutional repositories if and when these are set up. Significantly, the existence of The Depot means that all UK researchers in HEIs now have access to an OA repository in which they can deposit their work.

Secondly, once repositories are available, they need to be populated. At present, repositories typically house small amounts of material. This is changing, but slowly. Current moves by research funders have the potential to accelerate the population of repositories with research papers. An increasing number of research funders are mandating OA for output associated with their research. In the UK, several of the government-sponsored research funders as well as Wellcome have recently introduced OA mandates which are likely to make a major impact over a four-to-five-year period.

Thirdly, new services need to be set up to make repository content more accessible, usable and reusable. Most content in repositories can already be searched using standard web search engines, such as Google, although the way Google picks up repository content can be a little inconsistent. In addition, search engines provided by OAI service providers, such as BASE, are beginning to become more reliable. Some service providers are also developing value-added services which provide a variety of facilities to make access to repository content easier and more precise. The European Union-funded DRIVER project is investigating providing such an infrastructure for European research. There is potential for a wide range of innovative re-use services including text and data mining in this area.
Development of version identification standards

Once a large amount of material is deposited in repositories (Scenarios 2, 3 and 4), it will become increasingly necessary for clear version identification to be in place. This is particularly true since different versions of the same work may be available in different places. Preprints and postprints need to be labelled clearly for readers. In a broader sense, within academic communities there is a need to develop clearly understandable and accepted quality indicators in online environments.

Work is now ongoing to develop international standards for version identification. This is a complex project which needs to take into consideration a wide set of technical, procedural, managerial and cultural issues. A study by Sally Rumsey et al\textsuperscript{10} outlines many of the key issues.

The idea of several versions of papers coexisting in the public domain tends to undermine the traditional notion of the ‘definitive version’ of an article. Connected with this, the idea of ‘publication’ itself may be changing. Rather than it being seen as a single event, it may become more of a process, in which various iterations of a work are made public and where the work evolves in line with ongoing work by the author and response to peer comments.

Development of value-added features

OA repositories were originally conceived by OA advocates as vehicles for increasing immediate access to research papers. Papers, it was argued, should be made available in a ‘raw’ form simply to increase the access to and impact of research outputs. This is the assumption which underpins Scenario 2 above. However, the potential of repositories to deliver additional value-added services is currently becoming clearer. PMC is an example of a subject-based repository which is developing highly sophisticated value-added services. Some of these are based on data and text mining techniques. For example, pieces of data such as chemical structures can be identified in different articles and made searchable across the repository. More sophisticated subject searching may be possible using such techniques, which might also be implemented across different OA repositories as well as within a single repository. Services such as this are considerably easier to develop in an OA environment than one where access to outputs (either to humans or machines) is restricted. As more value-added features are developed in repositories, Scenarios 3 and 4 become more realistic possibilities.

At the same time, journals are now developing significant value-added features which may help to secure their long-term future. Some of these may involve interaction with repositories, and assume the coexistence of the two.

New business models

Scenarios 3 and 4 require the adoption of new business models associated with scholarly communication. The most common alternative model to the subscription-based system for journals is the publication-charge model. However, one major difficulty associated with this is that of transition: how can journals operating within one model move to another without substantially disturbing the scholarly communications system? David Prosser\textsuperscript{11} and Stephen Pinfield\textsuperscript{12} have suggested that the hybrid business model may provide a basis for a transition process.

The publication-charge or even hybrid model still present major procedural problems at present. Funding streams flowing from government agencies and research sponsors to research organizations or institutions cannot easily be ear-
marked for payment of publication charges. In the current system they are normally directed to funding subscriptions. Authors in institutions often have to go to a lot of trouble to ensure that publication charges can be paid. It will require major shifts at different levels of the system in order to create opportunities in the institutions for authors to pay publication charges easily. The University of Nottingham has recently tried to address this by setting up a central institutional fund against which members of the institution can claim OA charges. Setting this up took several months of discussions between Information Services and the research support office before it was ultimately authorized by the institution’s Research Committee.

Another challenge springs from the fact that at this stage it is not entirely clear what the costing and funding models will look like for repositories themselves. Institutions and subject communities have tended to make use of various funding streams to set up and maintain repositories, often relying on pre-existing technical infrastructures and in-house expertise. Whilst maintaining repositories is unlikely to present significant problems to these stakeholders, the cost should perhaps be surfaced more explicitly and highlighted as part of the cost of research itself.

More work needs to be carried out on the costs of scholarly communication and where these costs are borne in the process, which would inform an understanding of how different business models can be developed. Of course, many of the current costs for the process as a whole are hidden. This applies most notably to the costs of content creation and many of the costs of quality control (authoring and peer-reviewing articles) which are currently subsumed within the general costs of research institutions. Most alternative business models assume that such hidden costs will remain hidden. Any model that includes payments for peer reviewers, for example, would considerably raise costs for the system as a whole.

New approaches to quality control
The current system of scholarly communication has peer review at its centre. Very few accounts of how the scholarly communication system might change question the position of peer review, although many seek to improve the way in which it is administered. However, there are a number of ways in which peer review might be complemented by other quality-assessment processes.

Key post-publication quality measures may become more important and may, for example, be taken into account by funding agencies wishing to allocate research funding. Metrics such as usage and citation levels are potentially easier to create in an online and OA environment. There is clearly a need to develop standards for measuring usage and citation at an article level (rather than just journal level) which can easily be applied and widely understood.

Post-publication peer review may also become more important. Initiatives to promote so-called ‘open peer review’, where an article is put into the public domain and then readers are invited to provide peer-review-based commentaries, have been variable in their success. Some repositories which include preprints, such as arXiv, already facilitate something like this informally. Authors report that they sometimes receive evaluative comments on their preprints which enable them to revise the paper before submitting it for formal peer review. However, journals such as *Nature*, have also experimented with a more formal post-publication peer review but report a lack of engagement with readers.

Another form of post-publication peer review which has been put forward as a possible future model is the ‘overlay’ or ‘virtual’ journal. The overlay journal is one that is created using content already in the public domain (perhaps in
repositories) which is identified, selected and peer-reviewed by the journal. Articles are quality-assessed, appropriately labelled and then brought together in virtual journal 'issues'. However, there are as yet few working examples of this model and it presents a number of practical challenges. One key question is the extent to which peer review should be expected to lead to alterations to the content. Pre-publication peer review can often give rise to significant changes to a paper. It is not clear how this could be facilitated by overlay journals.

However, one major feature of this model remains important: the separation of quality control on the one hand from the publication process on the other. The current journal publishing system has the two closely tied together: quality control is managed by an editor and editorial board who oversee the work of peer reviewers and who publish quality-audited papers in a particular journal. It is possible that all these players could carry out similar roles outside a journal title. Peer-review could become a kind of ‘kitemark’ applied to papers which are actually made publicly available in various ways and in various places. Such a process would almost certainly be required if Scenario 4 above is to operate efficiently.

Adoption of digital preservation as a repository function
Scenario 4 also requires preservation of digital content to be closely associated with repositories. Currently, most repositories are seen as primarily providing immediate access to content and do not provide robust long-term digital preservation functionality. However, neither do journals in the current system. Digital preservation is still not adequately managed for a large proportion of online content wherever it is housed. More work clearly needs to be done in this area. Work such as this has been carried out by the SHERPA DP project, which has looked at a central preservation agency providing preservation services for a number of distributed repositories. This is one possible model of how preservation might be carried out in the future for distributed repositories. However, a great deal needs to be done to clarify technical, managerial and economic questions in this area.

Conclusion
Open access repositories and peer-reviewed journals are not necessarily in competition. There is a significant body of empirical evidence which shows that the two can (and do) coexist and may even be complementary. However, there is also evidence of uncertainty amongst major stakeholders in the field about the future. This means that an evidence-based dialogue between players should be promoted going forward in order to ensure that changes which do occur in the scholarly communication system benefit the research community.

There are a number of possible models of how scholarly communication may develop in the future, some of which involve repositories and journals coexisting. Whilst it is not clear precisely how the environment will end up, it is clear that today’s is a rapidly changing landscape. Both journals and repositories may both have an ongoing place in this landscape but if they are to continue to coexist in the long term, both need to change in significant ways.
References

4. Ware, M, ALPSP Survey of Librarians on Factors in Journal Cancellation, 2006, Worthing, ALPSP
13. See: http://www.nature.com/nature/peerreview/debate/index.html

*Article © Stephen Pinfield*