Short title for running heads:
A mandate to self archive?
Stephen Pinfield

A mandate to self archive? The role of open access institutional repositories

Based on a paper presented at the UKSG seminar 'Scientific Publications: Free for all?', The Geological Society, London, Tuesday 23 November 2004

STEPHEN PINFIELD
Director of Research and Learning Resources
Information Services
University of Nottingham

Abstract

This paper argues that the best way to achieve major improvements in scholarly communication in the short and medium term is to make it mandatory to deposit research papers in open access institutional repositories. This is what the House of Commons Science and Technology Committee report of 2004 [1] on scientific publishing recommended. The paper defines what open access repositories are and explains why they should be institutional. It also deals with question of what should be deposited in institutional repositories and why these improve scholarly communication. It then deals with the issue of mandating deposition: why deposition should be mandatory, who should mandate deposition and who should carry out deposition. The paper concludes with an analysis of the wider implications of mandating deposition in institutional repositories and a summary of the existing situation in the UK and elsewhere. The paper discusses the Select Committee report and the UK Government response in relation to institutional repositories.

Introduction

This paper argues that the best way to achieve major improvements in scholarly communication in the short and medium term is to make it mandatory to deposit research papers in open access institutional repositories. Of course, this is exactly what the Select Committee report recommended. Recommendation 44 states: "We recommend that the Research Councils and other Government funders mandate their funded researchers to deposit a copy of all their articles in their institution's repository... as a condition of grant..." This was one of at least 16 recommendations concentrating on institutional repositories in the report (recommendations 7, 42-48, 50, 52-56, 58 and 75).

The fact that the report gives institutional repositories such prominence is in itself worth noting. The accusation has been made that the Committee pre-judged the issues and that the inquiry was 'a solution searching for a problem'. However, the importance attached to institutional repositories by the Committee, far from being pre-determined, emerged during the course of the inquiry in response to the evidence. It was not there at the beginning. The original remit of the inquiry published in December 2003 did not mention institutional repositories at all. In contrast, another issue, scientific fraud and malpractice, which did feature in the remit, is dealt with only briefly in the report itself. On the issue of malpractice, the Committee again seems to have responded to the evidence it was given, and in this case concluded that there was little that needed changing. In both cases (institutional repositories and malpractice) the assertion that the Committee pre-judged the issues seems to be at odds with the facts.

Because the report was a thorough investigation of the issues which took into account the relevant evidence, it makes it all the more disappointing that the Government response was so non-committal. The response was put together by the Department of Trade and Industry and was clearly heavily influenced by the publisher lobby – even some of the phraseology shows this. It repeats several times that the Government wishes to create 'a level playing field' for all of the players in the scientific publishing market but it says little about how it proposes to do this. The Select Committee report demonstrates that a level playing field does not exist at the moment but, significantly, this is not acknowledged in the Government response.

In the area of institutional repositories, the Government response "recognizes the potential benefit of institutional repositories and sees them as a significant development worthy of encouragement" (page 27). However, it sees this as a matter which can be left entirely in the hands of institutions and it stops short of a mandate: "the Government has no present intention to mandate..." (page 28). Nevertheless, the Government does say that it is content to allow the work currently being undertaken by the Joint Information Systems Committee (JISC) and Research Councils UK (RCUK) to continue. Since RCUK is looking seriously at the possibility of a mandate, and the JISC has done a great deal to encourage innovation in scholarly communication, this is significant.

Key questions

The argument that 'the best way to achieve major improvements in scholarly communication in the short and medium term is to make it mandatory to deposit research papers in open access institutional repositories' needs further explanation. It begs a number of questions:

- What are 'open access repositories'?
- Why 'institutional' repositories?
- What should be deposited in them?
- Why do they 'improve' scholarly communication?

- Why make deposition mandatory?
- Who should mandate deposition?
- Who should do the depositing?
- What would happen then?
- What happens now?

These questions will be addressed in turn in order to demonstrate that deposition in the open access repositories, or 'self archiving' as it is sometimes known, has enormous potential.

What are open access repositories?
'Open access' needs defining first. There are a number of different definitions of 'open access' but most of them have key features in common. Open access exists where there is free, immediate and unrestricted availability of content. Some definitions also specify unrestricted re-use of the content but this is perhaps unnecessarily prescriptive. An open access *repository* is then an online database on the Internet which makes the full text of items (or complete files) it contains freely and immediately available without any access restrictions.

Why 'institutional' repositories?
The Select Committee report defined *institutional* repositories as "online archives set up and managed by research institutions to house articles published by authors at those institutions" (paragraph 108). Why, though, should repositories be institutional? Most advocates of institutional repositories would give a pragmatic answer to this question. Institutions have the technical and organizational infrastructures, the resources, and the expertise to set up and maintain repositories in the long term. They also have every reason to do so. Repositories can enhance an institution's profile and also help it to manage institutional information assets more effectively (to facilitate such activities as submission to the Research Assessment Exercise). In other words, institutions (such as universities) are in the best position to set up, maintain and populate repositories.

An institutional (or any distributed) approach to repositories is only workable in a context of interoperability. Repository interoperability is achieved by the Open Archives Initiative Protocol for Metadata Harvesting (OAI PMH). This technology creates the potential to expose metadata about the contents of a repository on the Internet so that it can be harvested. Metadata gathered from a number of different repositories can be collected into a searchable database by a third party. End users can then search harvested metadata from a single OAI search facility (or even Google). This means that the actual location of the full text itself does not matter to the end user.

The functionality associated with OAI PMH means that institutional and subject-based repositories can coexist and complement each other. The 'institutional versus subject' repositories debate is a red herring. Both can work at the same time. However, it needs emphasizing that distributed repositories alone are not enough. To improve scholarly communication what

is required is institutional (and other) repositories *combined with* effective search services and subject-based aggregators. The repositories have the content, while the search engines and aggregators provide user-friendly ways into it. With the technology already in place to achieve this, it is perhaps only a matter of time before significant numbers of repositories *and* search services begin to spring up.

What should be deposited in institutional repositories?

A wide variety of digital objects can be deposited in institutional repositories. Alongside research papers, data and other non-textual files can also be stored and made openly accessible. However, electronic versions of research papers, or 'e-prints', are at the heart of the current debate. An e-print is "a digital duplicate of an academic research paper that is made available online as a way of improving access to the paper." [2] The paper might be a 'preprint' (the version of the paper before it has been refereed) or are 'postprint' (the version that has been changed in response to referees' comments). It may be a book chapter, conference paper or similar research output, whatever is the norm in any given subject community.

The idea of depositing e-prints in institutional repositories raises a number of important practical questions. Firstly, which file should be used? Most e-print repositories will normally contain the author-produced files (in a format such as PDF). Some publishers (though not many) will allow *their* PDF (the version which has been copy-edited and formatted by them) to appear in an e-print repository. Where this is not the case, there is general acknowledgement that the author-produced e-print does not take on the role of the version of record (this continues to be provided by the journal). A second major issue is copyright. This was highlighted by the Select Committee report which recommended that in some cases author copyright retention (as opposed to copyright transfer to publishers) may be necessary in order to ensure institutional repositories can be populated. However, at present the majority of large publishers allow e-prints to be deposited in institutional repositories, and so although change may be necessary in the future, there is a great deal that can be done now. The third issue is quality. There is of course no reason why in principle high quality cannot be maintained in an open access environment. What open access outlets need to do now is not so much develop new mechanisms for maintaining quality but rather find new ways of flagging quality. Institutional repositories in particular (as the Select Committee report pointed out) need to develop clear quality markers ("kitemarks") that users recognize and trust.

Why do institutional repositories improve scholarly communication?

Whilst there are a number of issues that still need resolving in relation to institutional repositories, the benefits are clear. Institutional repositories improve dissemination of content – making it quick, easy, wide and cheap. They break down access barriers to content inherent in the subscription-based publishing system. The benefits of making scholarly content openly available in a timely way to anyone with a web browser are profound. Following this, once the content is easily available, interesting things can then be done with it. Search services can be developed using OAI PMH

technology – creating the potential for a global virtual research archive which can be searched from a single access point. The literature can also be analysed more easily. Text mining technologies can be implemented more effectively in an open access environment. Citation analysis at the article level can be carried out. Automated plagiarism detection can be implemented on a wide scale. All of these are very difficult to operate across different subscription-based services with access toll gates.

Open access also creates greater impact potential for research papers. The evidence for this is in two strands. Firstly, there is direct evidence that making a paper available on open access tends to produce more citations. Work has been done on a number of different disciplines and the evidence shows consistently that open access means more citations. Secondly, there is indirect evidence in this area. Open access usually creates more downloads – more readings of the article. When this fact is added to the second finding that downloads correlate closely with subsequent citations, these two findings together create another important strand of evidence that open access papers have a greater impact. In short, open access improves communication. It improves access to papers and improves the impact of papers.

These benefits are not just theoretical ones. They are already there to see (albeit currently in a limited way). The benefits of open access repositories can be demonstrated by looking at arxiv.org, which has been in existence for 14 years and has become indispensable to the high energy physics community. There is no reason to believe that these benefits could not be extended to other disciplines, even bearing in mind discipline differences.

In view of the benefits of open access, an increasing number of researchers are recognizing that they have every reason to self archive their work. One might go further and say that they have a 'mandate' to do so (using the word in a rather different sense). They give their work away for free, they participate in quality control activity for free (carrying out peer review and sitting on editorial boards), they want their work to be available for free and self archiving is a way of achieving this. This is a growing view but open access enthusiasts are still in a minority. It would take a number of years to change this.

Why make deposition mandatory?

This leads to the question: why make deposition mandatory (returning to the previous use of 'mandate')? If the benefits of open access are so clear, why 'force' the issue? The argument here is that mandating open access is the *best* way to improve scholarly communication but it is certainly not the *only* way. Making it mandatory would help to accelerate change and make the benefits more apparent across all subject disciplines, but there is an argument that this would happen anyway, without a mandate, given time. A mandate would simply help to overcome quickly the cultural and managerial barriers that currently exist in this area; something that would otherwise take a number of years.

Of course, the concept of ‘mandate’ does carry with it its own cultural problems. Many academic researchers do not like to be ‘forced’ to do anything. Nevertheless, research funders, institutions and other agencies already do ‘require’ researchers to do certain things sometimes ‘as a condition of grant’ (such as produce research reports or carry out certain administrative tasks). Requiring authors to deposit a copy of an e-print in a repository is in practical terms very little to ask (ten minutes’ work), particularly if they are given practical support.

Who should mandate deposition?

It might be institutions that require authors to deposit their papers. This is already being done by one or two institutions. Particular departments or schools *within* institutions might also introduce a local mandate. However, it perhaps makes most sense (as the Select Committee suggested) for research funders to introduce a mandate as a condition of grant. In the UK, the research councils (who fund most of the research in UK universities) could do this relatively easily. Other funders, such as the Wellcome Trust in the UK and the National Institutes of Health in the US, are leading the way. They recognize that open access repositories have the potential to make major and immediate improvements in the scholarly communication and, as a result, are moving in the direction of a mandate.

But should there be any delay in introducing such policies? Can all institutions be reasonably expected to comply in the short term? The technical barriers to entry are low. Setting up an institutional repository is well within the reach of most research institutions. There is free software to do it and a pre-existing network of institutional repository managers who already support each other. For those institutions which feel they cannot go it alone, there may be consortial options. There are also now commercial providers who will set up and run a repository for an institution. All of these options (institutional in-house, consortial, or commercial provider) are relatively low cost. Ideally, developments in the UK could do with co-ordinating, as the Select Committee recommended, and there are already in existence agencies (such as the JISC and the Research Libraries Network) that could carry out his role without delay. When this pre-existing organizational infrastructure is combined with the low technical and financial barriers, they create an environment in which there need be no significant delay in moving ahead with a mandate.

Who should do the depositing?

Institutions would have to set up internal support procedures to facilitate deposition. Authors might ‘self archive’ papers personally but it might also be possible for deposition to be done on their behalf by support staff in schools or central support services (such as the library). Once again, there are pre-existing structures that could easily support this activity in most if not all institutions. It would not take long to develop new policies to support large-scale content deposition in institutions.

What will happen then?

Once a large part of the literature becomes available via open access repositories (or other open access services) the benefits for the scholarly community would quickly become apparent. Open access improves scholarly communication. And since communication is the lifeblood of science and scholarship, better communication leads to better science and better scholarship. This is the key argument in favour of open access.

But there are other benefits. Society as a whole could benefit. Open access could lead to a better public understanding of science, better knowledge transfer between research institutions and industry, better dissemination of high quality content to a inform clinical practice. Making publicly funded research publicly available is likely to lead to considerable benefits for many parts of society. A certain amount of restructuring in the publishing industry and research libraries which would probably result from this seems a small price to pay.

What happens now?

The open access movement is gathering momentum. In the UK, the Select Committee inquiry has done a great deal to bring the issues in front of a broad range of stakeholders. The ongoing debate between the Government and the Select Committee has kept the topic in the public eye. Agencies such as the JISC and RCUK are continuing their work. Funders, such as the Wellcome Trust, continue to have a major input in the debate. The mandating of deposition by institutions or funders of research papers in open access repositories (institutional or otherwise) remains a real possibility.

On an international level, open access seems to be gaining more supporters. In Europe, an increasing number of institutions and funders are signalling their support for open access principles and their intention to work towards achieving greater open access in practice. In the USA, there is a growing open-access movement finding support amongst researchers, librarians and policymakers. Throughout the world, the open access movement is finding supporters.

Conclusion

It remains to be seen how these developments take shape but it seems likely that open access repositories in general and institutional repositories in particular are likely to play an important role in the future. They could certainly create major improvements in scholarly communication in a short time if they held a large proportion of the research literature. The best way to make this happen quickly and widely is to make deposition mandatory. It can only be hoped that key stakeholders and policy-makers have the vision and courage to see the opportunity and make it happen.

References

1. Great Britain, House of Commons, Science and Technology Committee, *Scientific Publications: Free for all? Tenth Report of Session 2003-04 HC 399-I & II*, 2004, London, The Stationery Office Limited ISBN 0215018419 for HC 399-I, 0215018419 for HC 399-II

web reference:
http://www.publications.parliament.uk/pa/cm200304/cmselect/cmsctech/399/399.pdf

2. Swan, A. *et al*, Developing a model for e-prints and open access journal content in UK further and higher education, *Learned Publishing,* 2005, 18(1), 25-40.

Stephen Pinfield
Director of Research and Learning Resources
Information Services
Hallward Library
University Park
University of Nottingham
Nottingham, NG7 2RD, UK
Tel: +44 (0) 115 951 5109
Fax: +44 (0) 115 951 4558
E-mail: Stephen.Pinfield@Nottingham.ac.uk
Web http://www.nottingham.ac.uk/~uazsjp/