

**Evidence on Scientific Publications for the
Science and Technology Committee of the UK Parliament
from the SHERPA Project**

Contents

- 1. Introduction**
- 2. Background**
- 3. Problems with the current system of scientific publishing**
- 4. Solutions: open access**
- 5. Key issues in the development of open-access institutional repositories**
- 6. The role of government**
- 7. Recommendations**
- 8. Appendices**

1. Introduction

- 1.1. The current system of scientific publishing has some key strengths. One strength is the rigour of the scientific process itself and the professionalism of those who practise it. Another strength is the robust system of quality control that has been developed. Quality is assured by peer review – the process by which scientists scrutinise each other's work before publication. But despite these strengths, there are a number of structural problems in the existing system of scientific publishing which mean that it is currently operating at a sub-optimal level. The evidence presented in this document briefly outlines what the problems are and discusses some of the ways in which they can be addressed whilst at the same time maintaining the system's key strengths. It suggests that scientific publishing should take full advantage of the opportunities provided by the World Wide Web by making high-quality research freely available online.
- 1.2. Providing free and unrestricted access (so-called 'open-access') to scientific papers has the potential to make scientific publishing more effective and efficient. Supporters of this approach advocate two complementary strategies – open-access journals and open-access repositories. Both strategies have realistic business models associated with them and have already begun to be implemented in various ways. This document concentrates on the case for open-access repositories. These repositories can be set up by individual institutions to store and disseminate the published work of their academic staff. This approach maintains the strengths of the current system, while making research available to the widest possible audience.
- 1.3. This submission makes 3 recommendations: (1) The setting up of open-access institutional repositories should continue to be encouraged within universities and other appropriate research organisations. (2) Authors should be discouraged from signing over exclusive rights to publishers and should retain (at least) electronic distribution rights for their papers. Where research work is publicly funded, it should be made a condition of grant that authors cannot sign-over copyright. (3) Public funding agencies should mandate (as a

condition of grant) that work they have funded must be made publicly available either in open-access journals or in open-access repositories.

- 1.4. This document is divided into a number of sections. After this introduction, section 2 provides details of the perspective from which this document has been written. Section 3 outlines some of the problems associated with the current system of scientific publishing. Section 4 discusses possible solutions to these problems, concentrating on the idea of open-access repositories. Section 5 provides more detail on how these repositories operate. Section 6 outlines the role government might play in developments, and this is followed by specific recommendations in Section 7. Background information is given in the appendices.

2. Background

- 2.1. The call for evidence states that the Science and Technology Committee “will be asking what measures are being taken in government, the publishing industry and academic institutions to ensure that researchers, teachers and students have access to the publications they need in order to carry out their work effectively?” This submission addresses that question.
- 2.2. This evidence is submitted by the SHERPA project. SHERPA (Securing Hybrid Environment for Research Preservation and Access) is a development project investigating the future of research communication and publication. In particular, it is initiating the development of openly accessible institutional digital repositories of research output in a number of UK research-led universities. These so-called ‘e-print archives’ will contain published (and in some cases pre-publication) research papers by academics from the participating institutions and will be made freely available on the World Wide Web. SHERPA is funded by the Joint Information Systems Committee (JISC) and the Consortium of University Research Libraries (CURL).
- 2.3. Because of the remit of the SHERPA project, the evidence presented here is limited to discussion of institutional e-print repositories. We understand that other organisations, such as CURL and SCONUL (Society of College, National and University Libraries) will be submitting evidence addressing all of the questions raised by the Inquiry. Institutional e-print repositories are part of a bigger picture of the possible future of scientific publication and we suggest a very important part. We believe that a strategy to reform research publication incorporating such repositories has the potential to make major improvements in the efficiency and effectiveness of the existing system.

3. Problems with the current system of scientific publishing

- 3.1. There are a number of problems associated with the existing system of scientific publishing which means it operates inefficiently and is getting worse. There are problems for researchers, research institutions and society as a whole. These are discussed in turn below.
- 3.2. Researchers communicate their research results in a number of ways, the most important of which is the peer-reviewed journal. In their capacity as authors, researchers give their articles to journals in order to achieve ‘impact’ not

income. They want to be influential in their field so that their work will be cited by colleagues. It is therefore in authors' interests that their work should be disseminated as widely as possible. However, publishers of journals want to generate income from research papers and so charge (substantial) subscription fees for journals whilst restricting their circulation (to subscribers only). Publishers also often require authors to sign over copyright completely, preventing authors from disseminating their work in other ways. This practice takes ownership of the work away from the researcher and the institution.

- 3.3. There are also problems for researchers in their capacity as readers of the journal literature. With over 24,000 refereed journals now being published¹, any one UK institution struggles to subscribe to a third of these. The effect is that most of the refereed literature is not easily available to most researchers. Such inefficiencies in the scientific communication process can hold back scientific progress itself.
- 3.4. Research institutions, of course, want to provide their members with access to as many journals as possible. This is essential if researchers in the institutions are to keep up-to-date in their field. But with growing lists of 'must have' journal titles, universities have for a long time been struggling to keep up subscriptions. Journal price inflation is a major problem. Between 1986 and 2000, journal price inflation in the UK was 291%, while the retail price index rose by only 74%². Even the development of e-journals has not helped to lower prices, in spite of reduced production costs for publishers. There is often a misunderstanding that all material available through the web is somehow free. In fact, publishers are attempting to establish e-journal-pricing models that maintain and increase their income. The 'Big Deal' is an example of this. Universities may experience the short-term gain of access to more titles. However, there are long-term problems of having to pay more for a bundle of titles only some of which are needed and where there is little flexibility to cancel individual titles.
- 3.5. From a corporate viewpoint, university institutions are currently in a position where they produce research output, then give it away free of charge to publishers, and then buy it back at high prices. Universities pay researchers to give away their services as authors and also as editorial board members, referees and journal editors (editors sometimes receive an honorarium from the publishers). This is in effect a massive subsidy to commercial publishers. Furthermore having handed over copyright to publishers, universities often have to buy back the content several times over: in journal subscriptions, photocopying licences, study pack charges etc.
- 3.6. The problems for researchers and their institutions create problems for society as a whole. Most academic research in the UK is funded from the public purse and yet the public does not have access to it. Furthermore, public money is being used to fund not just the research itself but also the buying back of that research by universities in the form of journal subscriptions. There is a strong argument which says that publicly-funded research should be made freely available to the public in general. The current system of restrictions on access to research output weakens the potential for knowledge transfer, acting as a

¹ As of August 2003. Source: analysis from Ulrich's Periodicals Directory, quoted in online discussion forum – <http://www.ecs.soton.ac.uk/~harnad/Hypermail/Amsci/2983.html>

² Swets Blackwell Periodicals Price Index, quoted in "Scholarly Communications Crisis" (briefing paper) Consortium of University Research Libraries (CURL), (2001)

brake to the take up of new technologies by industry. It also has the effect of hiding many of our cultural and intellectual assets behind commercial 'toll-gates'. The public understanding of science could be greatly enhanced by improving access to the scientific literature itself.

4. Solutions: open access

- 4.1. There is now an international movement aiming to address the problems associated with the current scientific publishing system. A key objective of this movement is that content should be made freely available wherever possible on the World Wide Web. The increasing ubiquity of the web has created a new opportunity to distribute content quickly and cheaply (without having to pay for production and distribution costs of printed paper). Such a strategy, it is argued, would increase the efficiency of scientific research communication and create wider economic and social benefits (by promoting knowledge transfer between the academic and commercial sectors and improving public access to science).
- 4.2. There is growing support world-wide for this 'open access' approach. A number of research funding agencies in the USA³ and Germany⁴ have, for example, recently released statements supporting open access. In the UK, the Wellcome Trust has expressed its strong support for open access⁵. Other funding agencies and research organisations are investigating the issues. There is also growing support at the ministerial level of world governments for the importance of open access to research, as shown at the recent Organisation for Economic Co-operation and Development (OECD) meeting where "Ministers recognised that fostering broader, open access to and wide use of research data will enhance the quality and productivity of science systems worldwide."⁶ Government action in the UK at this stage would undoubtedly help to increase the momentum for change.
- 4.3. The movement for open access has focused on two strategies: open-access journals and open-access repositories⁷. The two strategies are complementary and not competitive. Both achieve the key aim of improving access to the scientific literature. Both offer the potential for major efficiencies in scientific publishing.
- 4.4. Open-access journals are currently being set up from within the research community. These include *PLoS Biology*⁸ and *BioMed Central*⁹ journals. They continue to provide rigorous peer review but then make papers freely available on the web. They can afford to do this because they operate a new business model, one in which payment is shifted from the output to the input stage. In

³ The Bethesda Statement on Open Access Publishing, 20 June 2003 – <http://www.earlham.edu/~peters/fos/bethesda.htm>

⁴ Conference on Open Access to Knowledge in the Sciences and Humanities, 20 - 22 Oct 2003, Berlin – <http://www.zim.mpg.de/openaccess-berlin/berlindeclaration.html>

⁵ Wellcome Trust announcement 12 Nov 2003 – <http://www.wellcome.ac.uk/en/1/awtvispolpub.html>

⁶ See paragraph 17 of the document "Science, Technology and Innovation for the 21st Century. Meeting of the OECD Committee for Scientific and Technological Policy at Ministerial Level, 29-30 January 2004 - Final Communiqué" and Annex 1: "Declaration On Access To Research Data From Public Funding" - http://www.oecd.org/document/15/0,2340,en_2649_34487_25998799_1_1_1_1,00.html

⁷ See, for example, the Budapest Open Access Initiative – <http://www.soros.org/openaccess/>

⁸ PLoS Biology homepage – <http://www.plosbiology.org/>

⁹ BioMed Central homepage – <http://www.biomedcentral.com/>

other words, rather than readers (or their institutions) paying for content through subscriptions, instead authors (or their institutions) pay for publication. Such a strategy still means institutions (and their funders) pay for the publication process but the current evidence suggests that if this model was widely adopted they would pay considerably less whilst at the same time achieving the widest possible dissemination of content. We understand that evidence for these conclusions will be presented to the Committee by a number of open-access publishers. The evidence presented here will therefore concentrate on the second open-access strategy – open-access repositories.

- 4.5. Open-access repositories are a practical solution to the current problems in scientific publishing, which could have an immediate and positive effect on the scientific process. The number of open-access journals is still small compared with the number of subscription-journals, and they are likely to take some time to establish themselves. However, institutional repositories can be set up relatively quickly and can act as an effective and efficient means of distributing and preserving content.
- 4.6. At the heart of the open-access repository agenda is the view that researchers should be encouraged to continue to publish in peer-reviewed journals but at the same time put copies of their published papers in a repository. The repository makes the papers freely available to all at the point of access. As a result, institutional repositories can maintain the strengths of the existing system (quality control via peer review) but also create new benefits. The new benefits are that research can be made available to the widest possible audience (because of open-access) and also that it can be disseminated quickly (a paper could be mounted on a repository as soon as it was accepted for publication for a journal, for example).
- 4.7. Open-access repositories already exist. A number of subject communities have been using them for sometime and proved the utility and sustainability of the idea. The arXiv service for physicists has been running since 1991 and now has over 300,000 papers freely available online¹⁰. Many of these papers are high quality articles which have been published in peer-reviewed journals. Others are provisional version of papers not yet peer-reviewed (so-called 'pre-prints') but which it is still useful for researchers to see.
- 4.8. Since the development of open-access repositories was until recently limited to small numbers of subject communities, a number of institutions have set up repositories to encourage a wider range of disciplines to contribute to the open-access literature. In the USA, MIT¹¹ and Caltech¹² have been running repositories for about two years. In the UK, the SHERPA project is one of a number of open access institutional repository initiatives currently taking place. SHERPA (Securing a Hybrid Environment for Research Preservation and Access) includes 18, research-led universities and other research institutions as partners (a complete list is given in Appendix 1). We would recommend that the UK government and government agencies should continue to support the setting up of open-access institutions repositories in order to accelerate these developments.

¹⁰ arXiv homepage – <http://arxiv.org/>

¹¹ MIT digital DSpace repository – <https://dspace.mit.edu/>

¹² Caltech CODA repository – <http://library.caltech.edu/digital/>

- 4.9. One of the major strengths of the open-access repository movement is its technical infrastructure. Repositories can be easily set up in such a way to ensure they can talk to each other. When compliant with the Open Archives Initiative (OAI) protocol (details of which are given in Appendix 2) all of the different repositories world-wide can effectively form a single virtual research archive. The contents of the different repositories can be searched simultaneously by using special OAI Service Providers (of which there are now several) or even using mainstream Web Search Engines (such as Google).

5. Key issues in the development of open-access institutional repositories

- 5.1. A number of key issues are currently being addressed as part of institutional repository initiatives. These are discussed as follows: copyright and IPR, quality control and peer review, the relationship with journal publication, costs and sustainability, and preservation.
- 5.2. Copyright and Intellectual Property Rights (IPR) are an important and (sometimes) complex part of the debate currently surrounding scientific publishing. However, one thing is clear: there is no practical need for exclusive rights to be transferred to publishers in order for material to be published in their journal. Although the practice is still common, many journal publishers now do not require the transfer of exclusive rights, and some funding agencies do not permit it (for work they have funded). Many funding agencies in the US, for example, do not allow copyright to be signed-over to private companies. We would recommend that authors of scientific papers should be discouraged from signing over copyright to publishers and should at least retain the right to disseminate their paper online for non-commercial purposes (by, for example, depositing it in an institutional repository).
- 5.3. Since the taxpayer funds the majority of the research in UK institutions, government could kick-start open-access at the funding stage. Firstly, OST funding agencies could prevent the copyright of work they have funded being given away by researchers. Secondly, they could mandate (as a condition of grant) that copies of all papers generated from publicly-funded research should be made publicly available on open-access, such as in an institutional repository. We would recommend that these developments should take place. If implemented, they would have an immediate beneficial impact on scientific communication.
- 5.4. Any new developments should not be allowed to undermine the quality control of scientific literature. The most important form of quality control is currently peer review and there is no reason for this to change. The process of peer review is currently administered by journal publishers and carried out by researchers. This process is independent of institutional repositories. Different subject disciplines follow subtly different models of quality control and have different traditions of circulating work-in-progress. Institutional repositories can be structured to support any or all of these. Repositories can be set up to allow biomedical practitioners to only deposit peer-reviewed papers, whilst at the same time allowing physicists to deposit peer-reviewed papers and un-reviewed 'pre-prints' (a normal part of that discipline's communication culture).

- 5.5. Whilst journal publishers continue to administer the peer-review process, researchers should of course submit their papers to high-impact journals. The acceptance of a paper by a peer-reviewed journal is currently a well-understood 'quality flag'. Authors should continue to publish in such journals and can include details of the place of publication on the paper when it is deposited in an open-access repository.
- 5.6. There is no evidence to suggest that such a practice would destroy journals. On the contrary, the empirical evidence from the physics community shows that arXiv has not undermined journals. Physicists continue to submit their work to peer-reviewed journals as well as contribute to arXiv. Authors continue to value the quality control function the journals provide but also the rapid and wide dissemination that arXiv provides¹³.
- 5.7. The widespread adoption of open-access repositories may, however, over time change the role of journals. Traditional journal publishing bundles together peer review with distribution of content. These functions could, however, be unbundled. If open-access repositories increasingly become vehicles for content distribution, publishers could re-focus their missions to become managers of peer review, and also providers of other appropriate value-added services.
- 5.8. Scientific publishing will of course continue to cost money. Even in a model where access is "free", production is certainly not. Institutional repositories may, however, help to minimise production costs whilst at the same time maximising the impact of the research. Many of the key production costs are already borne by research institutions (and their funders) – carrying out the research, writing the papers, refereeing papers, providing editors and editorial board members. The key remaining cost is peer review (plus smaller costs for copy editing and document formatting). In the long-term, these also could be borne by institutions, perhaps at the input stage rather than as subscriptions (as in the open-access journal model discussed above).
- 5.9. In the short-term, the costs of setting up open-access repositories are minimal. Universities already have good IT infrastructures in place – local area networks which connect to the internet and widespread use of computer workstations. Given this provision, the connection cost and use of repositories is absorbed within existing overheads, so accessing the material is effectively free. Mounting a paper on a repository takes no more than 10 minutes, which is insignificant compared with the time spent on its creation. Maintaining a repository less significant in scale than maintaining an institutional web site.
- 5.10. Institutions will have an interest in maintaining repositories, a fact that will help to ensure their sustainability. An open-access repository can help to raise the profile and standing of an institution. From a practical point of view, a repository could also be used as an effective way of managing an institutions 'information assets' for submission of work as part of the research assessment exercise, consultancy work or other out-reach activities.

¹³ Stephen Pinfield, 'How do physicists use an e-print archive? Implications for institutional e-print services'. *D-Lib Magazine*, 7, 12, December 2001 – <http://www.dlib.org/dlib/december01/pinfield/12pinfield.html>

- 5.11. Institutions will also have an interest in preserving the digital content within repositories. The preservation of research material in the long term is something that currently lies outside the publication process *per se*. Long term access and preservation of articles and journals as material of record has traditionally been left to libraries. Preservation has always required continual investment in facilities for storage, curation and access and this has traditionally been handled by libraries as part of their base costs. As such, this is often a hidden and additional cost of the traditional publication process, which has been borne by research institutions. Digital preservation is likely to be more costly. The widespread adoption of open-access repositories offers the chance to make the preservation of material a fully integrated part of the scientific communication process for the first time, since institutional repositories would have responsibility for both distribution and preservation of content.
- 5.12. Institutional repositories have a greater potential than just distributors and preservers of scientific papers. There is also the potential to use them to store and provide access to scientific data and other related digital files. This would mean that a published scientific paper could sit alongside the data upon which it is based, giving other scientists access to the raw material of the research, as well as its published output. Such a scenario is very attractive to many scientists, although as yet not widely implemented. However, the success of projects such as the human genome initiative has illustrated that making scientific data available on open access has led to rapid scientific progress.

6. The role of government

- 6.1. The principles of the open-access initiative have growing support within the scientific community. However, there are often few real incentives for *individual* scientists to adopt this approach when many of the current reward-structures in science are tied closely to the traditional print-based system of communication. Scientists are often reluctant to depart from this without real personal incentives. Clear and active support from government and other nationally funded agencies is required to encourage the widespread adoption of open-access strategies. Such support would help to accelerate the transition process. Since the UK contributes a significant proportion of the scientific output of the international research community, progress in the UK would have a major international impact.

7. Recommendations

- 7.1. The setting up of open-access institutional repositories should continue to be encouraged within universities and other appropriate research organisations.
- 7.2. Authors should be discouraged from signing over exclusive rights to publishers and should retain (at least) electronic distribution rights for their papers. Where research work is publicly funded, it should be made a condition of grant that authors cannot sign-over copyright.
- 7.3. Public funding agencies should mandate (as a condition of grant) that work they have funded must be made publicly available either in open-access journals or in open-access repositories.

Further information and contact:
Bill Hubbard
SHERPA Project Manager
Hallward Library
University of Nottingham
NG7 2RD

tel: 0115 846 7657
fax: 0115 951 4558

email: bill.hubbard@nottingham.ac.uk
web: www.sherpa.ac.uk

February 2004

8. Appendices

Appendix 1 - The SHERPA Project

Appendix 2 - The Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)

Appendix 3 - A Selection of Existing Repositories and Service Providers

Appendix 1 - The SHERPA Project

SHERPA: Securing a Hybrid Environment for Research Preservation and Access.

SHERPA aims to investigate issues to do with the future of scholarly communication and publishing. In particular, it is initiating the development of openly accessible institutional digital repositories of research output in a number of research universities. These so-called 'e-print archives' will contain papers by researchers from the participating institutions.

The project partners now include the majority of members of the Consortium of University Research Libraries (CURL) who represent a significant proportion of the research-led universities in the UK. This shows a level of effort and commitment that will make a substantial contribution to the development of the e-print archive network.

The project will investigate the IPR, quality control and other key management issues associated with making the research literature freely available to the research community. It will also investigate technical questions, including interoperability between repositories and digital preservation of e-prints.

SHERPA Participants:

- Birkbeck College
- University of Birmingham
- University of Bristol
- University of Cambridge
- University of Durham
- University of Edinburgh
- University of Glasgow
- Imperial College
- Kings College
- University of Leeds
- University of Newcastle
- University of Nottingham
- University of Oxford
- Royal Holloway
- School of Oriental and African Studies
- University of Sheffield
- University College, London
- University of York
- British Library
- Arts and Humanities Data Service

SHERPA is funded by the Joint Information Systems Committee (JISC) and the Consortium of University Research Libraries (CURL).

It is hosted by the University of Nottingham.

For further information see the web site – <http://www.sherpa.ac.uk>

Appendix 2 - The Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)

Open-access repositories would be of limited value if each repository stood in isolation and had to be searched individually. One of the keys to the success of open-access repositories is that they can be searched as though they were one unified repository. This is achieved by use of the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). In order for the OAI-PMH to work each repository makes available a standard description of every item in its collection, like a series of virtual library index cards. These descriptions (known as "metadata") are then collected up by third-party "service providers", in a process known as "harvesting". This harvested metadata is then used as the basis for search services which search the metadata of all participating repositories.

Typically, a user will search for an item using the author's name, or the title of the paper, or search using keywords, associated with the article. The results of such searches identify an item and where it is held and make it accessible through a single mouse click. The experience for the user is a seamless one, thus giving the effect of a global virtual archive of research papers.

This means that a repository is very different from academics simply mounting their research on their personal homepages. The use of the OAI-PMH has the effect of standardising the way material is mounted, described, searched for and accessed, making research material far more visible and accessible.

There are therefore two types of participant in this process:

- Data Providers - the institutions that build and populate the repositories.
- Service Providers - the agents that harvest the metadata and provide search services.

In addition to search services, service providers can also offer a variety of added-value services. Examples of these services include citation records and analyses, downloading counts, etc.

The metadata for items in repositories can also be searched using normal search engines, like Google. However, one advantage of using dedicated service providers is that they need only search through registered academic repositories. This means that the dedicated search results from service providers are not cluttered with the thousands of junk results which are typical from a standard web search engine. This provides quick, clean and focussed access to research material held on open-access repositories.

Appendix 3 - A Selection of Existing Repositories and Service Providers

There are a number of existing repositories that are used by academic researchers. They can broadly be categorised as in two ways:

- subject-based, where the repository contains e-prints from a single subject-discipline, or a number of closely related subject-disciplines
- institutionally-based, where the repository contains e-prints produced by members of a single institution, typically covering a number of subject-disciplines.

Examples of Subject Based Repositories:

- Physics: arXiv – <http://arxiv.org/>
- Economics: rePec – <http://repec.org/>
- Cognitive Science: – CogPrints <http://cogprints.ecs.soton.ac.uk/>

Examples of Institutionally-based Repositories:

- University of Bath: "ePrints@Bath" – <http://eprints.bath.ac.uk/>
- University of Glasgow: "ePrints Service" – <http://eprints.lib.gla.ac.uk/>
- University of Nottingham: – "Nottingham ePrints"
<http://eprints.nottingham.ac.uk/>
- California Institute of Technology: – "Caltech CODA"
<http://library.caltech.edu/digital/>
- Massachusetts Institute of Technology: "DSpace" – <https://dspace.mit.edu/>

A select list of institutional repositories is available from the SPARC Europe website – <http://www.arl.org/sparc/core/index.asp?page=m1>

Examples of Service Providers:

- Arc, developed by the Old Dominion University – <http://arc.cs.odu.edu/>
- CiteBase, an experimental service developed by the University of Southampton which demonstrates some of the value-added features mentioned in Appendix 2 – <http://citebase.eprints.org/cgi-bin/search>
- OAlster, developed by the University of Michigan – <http://oaister.umdl.umich.edu/o/oaister/>