

Open Access to Scientific Information



The internet has transformed the nature of scientific research, opening up new ways to collect, use and disseminate scientific information. This has led to increased demand for access to such information. Open Access (OA) to scientific journal publications means making them freely available online, rather than charging readers to view them. OA to research data means making research data more widely available for re-use by others to support research, innovation and wider public use.

Background

The growth of the World Wide Web in the 1990s presented new opportunities for disseminating scientific publications. Scientific publishers have made substantial investments in online delivery and in digitising old content. Now, 95% of journals are available online. The Open Access (OA) movement seeks to make scientific (and other scholarly) publications freely available online. A drive for greater transparency and more effective exploitation of public funds has also led to demand for OA to publicly-funded research data, to support research and innovation.

Expanding access to scientific publications and data could deliver widespread social and economic benefits. There are several ways of achieving this. In March 2011 the Minister for Universities and Science held a round table discussion to explore the issues. At this meeting the government committed to supporting efforts to expand access to both research publications and data as part of its wider "Transparency Agenda". OA is seen by many as a key option for expanding access. However, there are challenges as well as benefits to providing OA to publications and research data. This briefing examines these separately before highlighting cross-cutting issues.

Overview

- Open Access (OA) to scientific publications could provide more effective dissemination of research and thus increase its impact.
- The costs and benefits of different models of providing OA to publications need to be considered if a comprehensive shift to OA is to be financially sustainable.
- OA to research data could enable others to validate findings and re-use data to advance knowledge and promote innovation.
- Sharing data openly requires effective data management and archiving. It also presents challenges relating to protecting intellectual property and privacy.
- Expanding access to scientific information requires researchers, librarians, higher education institutions, funding agencies and publishers, to continue to work together.

Open Access to Journal Publications

Scientific journals play a central role in the dissemination of scientific information, especially publicly-funded research. Journal publishers produce, distribute and archive printed and digital editions of journals. They also administer the peer review (PR) process, where articles are scrutinised free of charge by other researchers before publication.

The "Subscription Model" of Publication

Traditionally, publishers charge for their services through subscription fees to libraries and individual users. In the past couple of decades, journal subscription fees have, on average, increased above inflation, but library budgets have decreased, making access more of a challenge for publicly-funded libraries. Following the recommendations of the 2004 House of Commons Science and Technology Committee inquiry into scientific publications,¹ researchers, librarians, higher education institutes (HEI), funding agencies and publishers have been working together to improve access to publications. Although there have been significant advances, there is still much debate as to whether the subscription model is the most effective way to disseminate research for maximum access and impact.

Some researchers, funding bodies and HEIs have argued that the cost of some journal subscriptions is compromising

the UK's capacity for research. They say that researchers, especially in smaller UK universities, do not have access to the publications they need. Some have also argued that the subscription model does not deliver adequate access to researchers in countries with lower income levels, or to the wider public, such as individuals with medical conditions. However, many subscription publishers argue that the subscription model is highly competitive and efficient and that UK research libraries could make efficiency savings elsewhere, to accommodate subscription costs. They also say there are already several initiatives to expand access, such as provision of free or low cost access online access to developing countries.

In September 2011, the government set up an independent working group to look into how expanded access might best be achieved. This group aims to produce recommendations in spring 2012. Many researchers, funding bodies and HEIs argue that OA could offer a key way to expand access and increase the impact of research.

Why Open Access to Publications?

OA proponents argue that, as a matter of principle, the knowledge emerging from publicly funded research should be made freely available to society. They also argue that free online access to publications would lead to social and economic benefits, resulting from increased communication and translation of knowledge. A recent study suggests that a shift to OA could increase the return on R&D by between £184 and £386m over a 20 year period, depending on how OA is provided.² This reflects the increase in access provided to government, the NHS, corporations and small to medium enterprises, as well as universities.

There is a growing trend towards OA in the UK and worldwide. Around 20% of all articles were made available through some form of OA in 2009.³ An increasing number of UK research funding agencies including the Wellcome Trust and Research Councils UK (RCUK) and many HEIs have now adopted OA mandates. These require researchers to make their published articles freely available online. Most subscription publishers have incorporated OA models in their business to allow authors to meet funder and HEI requirements in ways that the publishers perceive are sustainable for the journal. The main challenge of OA is seen to be how to meet the costs of publication, particularly the administration of peer review, on a sustainable basis.

Models of Open Access to Publications

There are two main models of OA, and it is widely believed that there will be a mixture of the two alongside the subscription model for the foreseeable future.

- **“Green” OA:** the author deposits the "post-print" version of articles accepted for publication in an online repository which makes them freely available (see Box 1). Some subscription publishers specify embargo periods of 6-48 months, during which time the articles are available only from the journal. This is to protect their subscription income and secure a return on their investment.
- **“Gold” OA:** published articles are made immediately and freely available by the journal publisher in return for an

article processing charge (APC). A few publishers such as the Public Library of Science, BioMed Central and Hindawi provide OA to all of their journals. Many subscription publishers such as Elsevier and Nature Publishing Group have recently launched a few OA journals. Many subscription publishers offer “hybrid” models for some of their journals, under which authors can choose to make their articles OA by paying an APC.

Although OA can simply mean making articles free to read online, it usually also means removing some of the copyright restrictions on the reuse of that information, so that others can build on the published work. The 2011 Hargreaves Review on Intellectual Property (IP)⁴ reported that some aspects of current IP law relating to publications were obstructing innovation and economic growth. One example is the placing of restrictions on ‘text-’ and ‘data-mining’ where articles can be automatically searched and analysed. This can help to identify potential avenues for further research. With OA, such restrictions are removed.

Box 1. ‘Green’ Open Access Repositories.

Around 56% of publishers worldwide formally allow authors to self-archive the "post-print" version (the final draft after peer review, not including copy editing carried out by the journal).⁵ Repositories can also hold "pre-prints" (pre- peer review), research data, and other materials used outside the formal channels of publication. There are two main types of repository.

- **Institutional Repositories (IRs):** maintained locally by research institutes. These are seen as a good way to showcase and manage research. There are over 80 well established UK IRs including UCL, Cambridge and Southampton universities.⁶
- **Thematic Repositories (TRs):** usually organised by discipline; often holding a high proportion of international research. Examples are ArXiv for physics and computer science, UK PubMed Central for life sciences, and the ESRC Research Catalogue for social science.

The costs and benefits of building and maintaining repositories vary greatly. TRs tend to have formal business models with explicit costs met from funding agencies or HEIs. However the funding and benefits of IRs are generally not specifically quantified, with costs largely hidden within the public funding of HEIs. There is some debate as to how well some IRs are being used.⁷ IRs and TRs can be used in conjunction with each other, with IRs collecting all the content produced by the HEI, and TRs adding further index or search functionality at a discipline level.

Sustainability of “Green” and “Gold” Models

With green OA the costs of publication continue to be paid for through journal subscription fees, but anyone can access the publication via a repository, regardless of whether they subscribe to the journal. The cost to HEIs could rise slightly because of the need to fund repositories alongside paying for journal subscriptions. However, proponents say that this is outweighed by benefits to the UK through increased returns on R&D, as a result of wider access. Green OA could provide a cost-effective route to improving access, largely because digital repositories have already been built. However, to maximise the benefits, there is pressure on publishers to make embargo periods as short as possible. Some subscription publishers argue that this could also lead to significant levels of subscription cancellations, which could reduce revenues to the point at which some publications become non-viable. Many journals are

experiencing cancellations, although as yet these cancellations have not been clearly demonstrated to be due to OA repositories. EU funded research is attempting to shed light on this issue.⁸

Under gold OA publishing, the costs of publication are typically shifted from the reader to the author, in the form of an APC, which is usually paid by the author's funding agency or institution. This is seen by some publishers and by RCUK as providing a more financially sustainable business model compared with green OA. However, the one-off cost to the UK of a transition to gold OA is estimated to be in the region of £7 million, because of the need for HEIs to create central funds to meet APC charges, and for funders, HEIs and publishers to develop administrative systems to deal with APCs.² Some OA advocates argue that universities may not be able to provide the funds necessary to pay for the transition to gold OA unless they implement parallel cancellations in subscriptions.

There are also significant uncertainties about how gold OA will evolve and whether there would be a net decrease or increase in publishing costs in the UK under a wholly gold OA system. An increase would be more likely if the UK shifted to gold OA for its own funded research, but other countries did not. For universities the increased cost of paying APCs would need to be outweighed by a fall in subscription payments, as more articles became available as gold OA. HEIs would be expected to reduce journal subscriptions and divert the savings into paying APCs. However, savings are hard to calculate as there is significant variation in APCs between journals. The average APC is around £1,500, but they can be as high as £3,000.² A recent report suggests that if the average APC is under £2,000, costs to HEIs would stay the same and there would be significant economic and social benefits to the UK.² At the lower level there could be significant savings to HEIs, but at the higher level, although there is a net benefit to the UK in the returns to R&D, the cost to HEIs would rise significantly above the current subscription model.

Subscription publishers argue that APCs for high-status journals must reflect the costs associated with handling large numbers of rejected papers. However, some research funders, libraries and OA publishers argue that subscription publishers calculate APCs on the basis of current profit margins rather than costs incurred. Some members of the research community suggest that "hybrid" models can lead to "double dipping", whereby institutions are paying twice, for APCs and for subscriptions. However, subscription publishers argue that subscription costs will reduce as the proportion of gold OA articles grows.

Open Access to Research Data

There are various benefits from sharing research data more widely. Transparency and access to data enable errors or flawed research to be detected, and deter selective reporting of results. Recent high profile controversies, such as the unauthorised release of emails from the University of East Anglia's Climatic Research Unit in November 2009,⁹ have led to greater public demand for transparency and

access to research data, especially where it affects policy making. This is reflected in Freedom of Information (Fol) legislation, which has transformed the degree of openness required of scientists working in HEIs and public research institutes (Box 2). The routine sharing of some private sector research could also be in the public interest, such as data from clinical trials, and safety analyses by private companies, which are used to inform policy decisions (for example in the case of the 2011 oil spill in the Gulf of Mexico). Making data more widely available for re-use would also allow publicly funded research to be exploited more efficiently by the public and private sectors.

Most research funders, such as RCUK, have developed policies which promote access to research data. The Department for Business, Innovation and Skills has established a forum to consider the role of government in improving access to research data, to support research and innovation. However, although there are significant opportunities in sharing data, there are also substantial challenges. The Royal Society has recently launched a group to look at barriers to data sharing.¹⁰

Box 2. Legislation Covering Release of Research Information

- The Freedom of Information Act 2000/ Freedom of Information (Scotland) Act 2002 give public right of access to information held by public authorities.
- The Environmental Information Regulations 2004 / Environmental Information Regulations Scotland 2004 give the public right of access to environmental information held by public authorities.

"Information" includes research, meeting minutes and emails. Anyone can make a request and there is a presumption in favour of disclosure. Information can be withheld only if it falls under specified exemptions and the authority can demonstrate that it is not in the public interest to disclose it. Exemptions can apply to information that is confidential, commercially sensitive, or intended for future publication.

Making Data Open

Timing of Data Sharing and Intellectual Property

Many funding agencies and journals require authors to make all data that are necessary to reconstruct the process of analysis freely available at the time of publication. However, many funding agencies and HEIs argue that there is a balance to be struck between making data openly available for wider re-use, and the need to protect the research process and allow universities to exploit their intellectual property (IP). Because this balance varies across disciplines, funding agencies' requirements on when data should be shared vary greatly.

HEIs have expressed concern that Fol legislation could enforce data sharing at too early a stage, which could hinder research and commercial collaboration. They have raised concerns over the inclusion of research information in the Fol Act, alongside other information held by public authorities (Box 2). They say that there needs to be an exemption to protect ongoing research, as there is under the Scottish Fol Act. A backbench amendment to the Protection of Freedoms Bill (2010-11) seeks to amend the Fol Act to introduce an equivalent research exemption. However, the outcome of this will not be known until later in 2012. There have also been concerns over lack of clarity in the legislation regarding exemptions. After the

recommendations of the 2010 House of Commons Science and Technology Committee inquiry into the UEA Climatic Research Unit's e-mails,¹¹ the Information Commissioner's Office has published additional guidance on FoI legislation and research information for the higher education sector.

Data Protection

There are tensions between sharing data openly, especially in the biomedical and social sciences, and protecting privacy. The Data Protection Act 1998 (DPA) regulates the processing of "personal data" through restrictions on how the data can be recorded, stored, altered, used or disclosed. The DPA implements the EU Data Protection Directive 95/46/EC, which is being revised to take account of new technologies and the changing ways that personal data are being used. A draft instrument is due in late January 2012.

Prior to this revision, there had been major concerns over the effectiveness of the DPA and whether it fully reflected the EU Directive.¹² Under the DPA, personal data means data related to a living individual who can be identified, either directly or indirectly, from the data, or from other information held by the same organisation. In some situations the only legitimate way of using personal information for a purpose different from that for which it was originally collected, is by obtaining consent. However, if it has been anonymised (personal identifiers or codes removed) data protection legislation does not apply. However, it is possible to 're-identify' anonymous records through cross-correlation with other databases.¹³

The revision of the EU Directive provides an opportunity to address concerns. However, a balance is needed between protecting privacy, and ensuring that data remain useful for research and innovation. A report by policy think tank RAND says regulation should be based on general best practice principles, rather than stipulating how data are processed.¹⁴

Data Management and Archiving

To enable others to find and to re-use research data effectively, data need to be archived and made publicly accessible, with sufficient "metadata" (such as a full description of software programs) to allow others to understand the data. The Natural Environment Research Council and the ESRC have longstanding systems in place for sharing and archiving data. However, overall, the systematic retention and archiving of data and "metadata" are patchy. Changing practices could require a significant amount of time and money. To address this, funding agencies are introducing data management plans as an integral part of funding applications. Decisions also need to be made about which data have long-term value, and should be preserved in accordance with relevant standards, if they are to remain accessible and usable in the future.

Cross-Cutting Issues

The Role of Publishers

Subscription publishers can facilitate green OA by permitting authors to self-archive the "post-print" manuscript if they do not already do so. They can also archive manuscripts on the

author's behalf if the author agrees. They can promote gold OA through both the 'hybrid' model (by building a clear OA option into the manuscript submission process) and by launching new OA journals. Both subscription and OA publishers also have a role to play in ensuring data are made publicly and permanently available. However, there is agreement that it is not their role to manage large datasets, and that the wider research community should take the lead in data management.

The Role of Funding Agencies and HEIs

Although most funding agencies and HEIs now have policies regarding OA to publications and data, compliance is acknowledged to be patchy and some say processes need to be simplified for researchers. Funding agencies and HEIs could facilitate a shift to OA by communicating the benefits to researchers more effectively, ensuring policies are clearly underpinned by explicit budgetary and resource provision where necessary, and by monitoring compliance. Increasing the recognition of data management and sharing as a professional achievement (alongside publication record) will encourage researchers to invest their time in making their datasets publicly available for re-use. HEIs could also facilitate data sharing by linking data to the text of research reports made available through IRs (see Box 1).

The Role of Researchers

Access to funding is reported to be one of the main barriers to gold OA. However, increasing the uptake of existing "green" publisher policies, which are free to the researcher, and encouraging data sharing, will require a substantial shift in community attitudes and behaviour in some disciplines. Although funder and HEI mandates could increase OA, there is wide agreement that a persuasive rather than punitive approach is preferable, and the reward structures of academia could be adapted to incentivise researchers to adopt OA practices. The Higher Education funding bodies are considering how OA could feature within the new system for assessing the quality of research in the UK. However OA will not become a formal part of the Research Excellence Framework until 2020. Further, they say that a commitment to mandate specific forms of OA would have to be implemented in a way which is sympathetic to cultural variations within different research disciplines.

Endnotes

- 1 HC Science and Technology Committee Tenth Report of Session 2003-04, Jul 2004. *Scientific Publications - Free for All?* HC 399-1.
- 2 Cook *et al.*, Apr 2011. *Heading for the Open Road*, RIN.
- 3 Björk, B. *et al.*, 2010. OA to the Scientific Journal Lit. *PLoS One*, 5(6): e11273.
- 4 Hargreaves, I., 2011. *Digital Opportunity: A Review of Intellectual Property and Growth*. Intellectual Property Office.
- 5 Romeo registry of OA Policies, <http://www.sherpa.ac.uk/romeo/statistics>.
- 6 Registry of Open Access Repositories (ROAR), <http://roar.eprints.org>.
- 7 Friend, F, Sept 2011, *Open Access Business Models*, Knowledge Exchange.
- 8 <http://www.peerproject.eu/>
- 9 Russell *et al.*, Jul 2010. *The Independent Climate Change E-mails Review*.
- 10 Boulton *et al.*, 2011. Science as a Public Enterprise. *Lancet*, 377 (9778): 1633-35.
- 11 HC Science and Technology Committee First Report of Session 2010-11, Jan 2011. *The Reviews into the UEA's Climatic Research Unit's E-mails*. HC 444.
- 12 Liberty, 2007. *Surveillance and Personal Privacy in Modern Britain*.
- 13 Ohm, 2010. Broken Promises of Privacy. *UCLA Law Review*, Vol. 57.
- 14 Robinson, N. *et al.*, 2009, *Review of European Data Protection Directive*, RAND.