Alive and kicking: a progress report on Open Access, institutional repositories, and health information[*]

Peter Morgan, Medical Librarian, Cambridge University Library
pbm2@cam.ac.uk

Summary: The concept of Open Access has made substantial progress in some areas, not least that of biomedical research; but it has been slow to affect the health-care environment.

Introduction

Francis Bacon's dictum 'Money be like muck, not good except it be spread' might equally well be applied to information. The belief that information too needs to be spread more widely lies behind the Open Access (OA) movement and the emergence of institutional repositories. The concept of OA was formalised five years ago and has made substantial progress in some areas, not least that of biomedical research; but it has been slow to affect the health-care environment.

The Internet, the Web, and Google have changed our world fundamentally and for ever, giving rise to new patterns of social and professional behaviour, new rights, expectations, and obligations. Digital information raises issues of production, ownership, dissemination, use, and preservation. OA and repositories offer new ways of addressing these issues. They also have the potential to impact on the roles that health information professionals perform at different stages in the information cycle: as controllers (employers, administrators, and policy-makers); as creators of information (authors, editors, and commissioners); as intermediaries (knowledge managers, disseminators, publishers, trainers, and curators); and as end-users (practitioners and consumers).
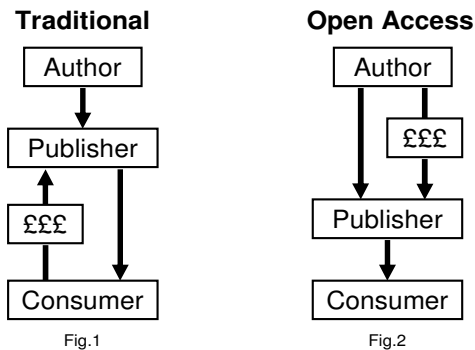
What is Open Access?

In February 2002, the Budapest Open Access Initiative (BOAI) <www.soros.org/openaccess/> defined Open Access as 'the world-wide electronic distribution of the peer-reviewed journal literature and completely free and unrestricted access to it by all scientists, scholars, teachers, students, and other curious minds', and other subsequent statements have promoted the same core principle.

The BOAI advocated two complementary strategies, OA publishing and self-archiving, to achieve its aims. OA publishing introduced an alternative to the traditional business model for journal publishing. In the traditional model (Fig.1), while information in the form of a paper flows from author to publisher to consumer, the process is funded by a financial transaction between the *consumer* (or the consumer's library) and the publisher, either as a subscription or as a pay-per-view

1

transaction. OA advocates argue that this imposes an unacceptable barrier between the information and its potential consumer, and in the new OA publishing model (Fig.2) the same information flow takes place but the process is funded by a financial transaction between the *author* (or the author's institution) and the publisher, thus removing the barrier between consumer and information.

Publishing models

Traditional

Author → Publisher
£££ → Publisher
Publisher → Consumer
£££ → Publisher

Fig.1

Open Access

Author → £££
Author → Publisher
£££ → Publisher
Publisher → Consumer

Fig.2

The OA publishing model has evolved into variant forms. Among the 'pure' OA journals the best-known, all in the biomedical field, are the PLoS (Public Library of Science) titles <www.plos.org/journals/index.html> and those published by BioMed Central <www.biomedcentral.com/browse/journals/>. At the same time, some traditional publishers have begun to develop OA options, either by publishing specific OA titles or through 'hybrid' journals which allow authors to choose between paying to have their papers published in an OA version or relying on publication through the consumer-pays model, all within the same journal title. Springer's 'Open Choice' <www.springer.com/west/home/open+choice/> and OUP's 'Oxford Open' <www.oxfordjournals.org/oxfordopen/> schemes have led the way, with other major publishers such as Blackwell, CUP and Elsevier following suite.

The second BOAI strand, self-archiving, has also made progress. It requires the author (or someone acting on the author's behalf) to deposit copies of pre-prints and/or published papers in a digital repository which is Internet-accessible and conforms to Open Archives Initiative (OAI) <www.openarchives.org/> metadata standards, and from which the papers are freely available to all. According to recent estimates, over 90% of publishers now permit some form of self-archiving within their general copyright transfer agreements. Authors wishing to check on publishers' self-archiving policies can refer to the searchable SHERPA RoMEO database of publisher copyright policies <www.sherpa.ac.uk/romeo.php>.

What are institutional repositories?

Open Access initiatives originated in the desire of many researchers to harness the potential of the Internet as a means of disseminating their papers more widely. Starting in 1991 with the creation of arXiv (physics), the early repository initiatives were subject-based. When it became apparent that repository content would be more readily retrievable if there was a common interoperability standard, the OAI's Protocol for Metadata Harvesting (OAI-PMH) was devised and has been built into new Open Source repository platforms such as EPrints (2000)

and DSpace (2002) <www.dspace.org/>. These have been widely adopted by universities and other research organisations, giving rise to institutional - as opposed to subject - repositories. For institutions lacking the resources to implement and develop Open Source platforms, commercial alternatives have also appeared, such as BioMed Central's 'Open Repository' <www.openrepository.com/>, based on DSpace, and ProQuest's 'Digital Commons' <http://umi.com/products_umi/digitalcommons/>, based on Berkeley Electronic Press's platform.

While the exact role and characteristics of institutional repositories may differ from one place to another, they fall broadly within the definition given by Crow in 2002[1] as 'digital collections capturing and preserving the intellectual output of a single or multi-university community'. Whether a repository may be termed 'Open Access', however, depends not on the platform but on the policies governing access to its content.

Whether institutional or subject-based, OA repositories will all have certain features in common: they capture, index, disseminate, and preserve digital content; they include content with a scholarly or professional purpose; they are cumulative, perpetual, secure; they observe interoperability standards and are cross-searchable by external search engines such as Google; they accept a variety of file formats; the files possess persistent identifiers; and, crucially, at least some of the content is openly accessible.

The progress of the OA movement can be illustrated with evidence from two directories. The scale of OA publishing is demonstrated by the Directory of Open Access Journals (DOAJ) <www.doaj.org/>: it currently (May 2007) lists 2,697 OA titles, of which 805 are indexed at article level. Similarly, the Directory of Open Access Repositories (OpenDOAR) <www.opendoar.org/> provides evidence of the progress made by OA repositories in acquiring content, listing 881 repositories in 46 countries.

But is OA working?

While the statistics suggest that progress is indeed taking place, some significant doubts about the future of OA remain. First, the OA business model is still unproven: many authors and their institutions appear reluctant to invest additional funds in this option, and pure OA publishers have therefore struggled to cover their costs. Secondly, the rate at which researchers self-archive their publications is disappointing: studies suggest that while up to 95% would submit their papers if it was a mandatory requirement, the number actually doing so on a voluntary basis is far lower[2]; and in the United States, where the National Institutes of Health in 2005 introduced a voluntary self-archiving policy, the rate of compliance in the first year was less than 4%. Thirdly, while most publishers now permit some form of self-archiving, many authors remain unaware of this and fail to utilise the advice obtainable from the RoMEO database.

It seems likely that self-archiving of papers will only become widespread if it is mandated by researchers' employers or funders. There was noteworthy activity in this respect during 2006 when five of the eight UK Research Councils introduced some form of mandatory deposit. Summaries of their policies can be consulted on the SHERPA JULIET database <www.sherpa.ac.uk/juliet/>.

It has also become apparent that many librarians, as repository managers, have taken a more holistic approach, developing repositories that contain both peer- and non-peer-reviewed literature as well as non-text materials; applying closed-access controls to some collections; and devoting significant resources to digital preservation issues. Some OA advocates ('archivangelists') argue that such practices have obscured the original 'pure' vision of repositories as a means of providing free open access to peer-reviewed literature.

Implications for health information

It has already been noted that the early front-runners in OA publishing were in the field of biomedical research. Of the 2,697 titles, 360 (~13%) listed by the DOAJ are in the field of health sciences (Dentistry 27, Medicine 231, Nursing 16, and Public Health 86) with another 148 titles in biology and life sciences (~6%). Similarly, among the 881 repositories listed by OpenDOAR, 72 (8%) have collections categorised specifically under the heading 'Health & Medicine' while more such material is distributed within general collections.

The mandates announced by major UK research funders are advancing the process further. Leading the way, the Wellcome Trust introduced its own mandate in October 2005, requiring deposit of all Trust-funded research papers in UK PubMed Central <http://ukpmc.ac.uk/>, a new purpose-built subject repository launched in January 2007. With the Medical Research Council and the Biotechnology and Biological Sciences Research Council adopting similar mandates, openly accessible biomedical research results promise to become a substantial resource.

It will be apparent from this review that most OA activity has so far occurred within the higher education research sector, leaving other sectors such as health care relatively unaffected. This situation appears to be changing, with signs that OA is being espoused to serve political agendas in health information, as seen in calls for clinical drug trial data to be made openly accessible[3], and in the launch of Open Medicine <www.openmedicine.ca/>, a new journal founded by staff who left the Canadian Medical Association Journal in an argument over editorial independence[4].

But for the NHS, perhaps the most significant development is the news that the Department of Health, already part of the eight-member consortium that funded UK PubMed Central, has introduced its own mandate[5] with effect from April 2007. This mandate requires deposit of papers supported by DH funding in a move that heralds a new era for dissemination of NHS research outputs and suggests official endorsement of Open Access as a principle to be embraced by the NHS more widely.

References
[all accessed 24/05/07]

1.Crow R. The case for institutional repositories: A SPARC position paper. SPARC, 2002. <www.arl.org/sparc/bm~doc/ir_final_release_102.pdf>

2. Swan A, Brown S. Open access self-archiving: an author study. Key Perspectives, 2005. <www.keyperspectives.co.uk/openaccessarchive/reports.html>

3. Adams M. The health care reform legislation that Congress should pass, but won't. News Target, 15 February 2007. <www.newstarget.com/021610.html>

4.Talaga T. Fired editors launch online medical journal. Toronto Star, 2 April 2007. <www.thestar.com/article/198412>

5. Statement on DH / NIHR-funded research and UK PubMed Central. <www.nihr.ac.uk/files/pdfs/OpenAccessPolicyStatement.pdf>