

Another reason for opening access to research

John Wilbanks

Science Commons Project, Creative Commons, c/o Massachusetts Institute of Technology Computer Science and Artificial Intelligence Laboratory, Cambridge, MA 02127, USA wilbanks@creativecommons.org John Wilbanks executive director

BMJ 2006;333:1306-8

Much of the debate about open access has focused on the principle of access for scientists and the economics of such a change. There is a second principle to consider: the full power of new technological approaches such as text mining, collaborative filtering, and semantic indexing, are not resulting in powerful new public resources. Despite real success in the open access movement, most scholarly research is unavailable, either for study or for processing by software.

According to the Budapest open access initiative, "There are many degrees and kinds of wider and easier access to this literature. By 'open access' to this literature, we mean its free availability on the public internet, permitting any users to read, download, copy, distribute, print, search, or link to the full texts of these articles, crawl them for indexing, pass them as data to software, or use them for any other lawful purpose, without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. The only constraint on reproduction and distribution, and the only role for copyright in this domain, should be to give authors control over the integrity of their work and the right to be properly acknowledged and cited."¹

Open access journals are entering the mainstream of scholarly publishing.² The Directory of Open Access Journals (www.doaj.org), a listing of "free, full text, quality controlled scientific and scholarly journals," includes 2478 journals, with on average more than one journal a day added in 2006 (121 999 articles are tracked).³ Open access journals have earned top impact factors in fields such as biology and bioinformatics,^{4 5} as well as high immediacy factors.⁶

But most journals are non-open access. Many non-open access journals and publishers tackle the access problem by allowing authors to post copies of their works on the internet; this is known as "self archiving."

Self archiving by authors is also growing rapidly. Between March 2005 and October 2006, the number of institutional archives tracked at the Registry of Open Access Repositories (<http://archives.eprints.org>) has grown by nearly one every other day, and the number of records in those archives has grown by nearly 600%, to 1.2 million papers. Open access is here to stay, in one form or another.

New business models, too

Much of the current debate around open access centres on economics and business models, which have been proliferating. As many as nine models have been described in studies of open access.¹⁴ Experimentation by new open access publishers and traditional publishers is sure to continue, with traditional publishers such as Springer and Oxford University Press experimenting with blended approaches.

Early returns on full open access are intriguing—studies show that open access at least has the potential for economic viability.^{15 16} On the negative side, a widely cited article in *Nature* questioned the viability of the non-profit model pursued by the Public Library of Science and noted increasing article charges at BioMed Central.¹⁷ However, Hindawi, a for-profit publisher with 49 open access journal, says its open access collection is already profitable.



Fig 1 The collaborative filtering and tagging site Connotea

Open access is important to the advancement of science for several reasons. Mostly the debate has centred around the arguments of price and access: "Authors communicate with only those of their peers lucky enough to be at an institution that can afford to purchase or license access to their work. Readers have access to only a fraction of the relevant literature, potentially missing vital papers in their fields" (see www.createchange.org).

Barriers of price and access have a particular impact on researchers in developing countries and at less wealthy institutions in developed countries. The impact of this is obvious: given the complexity of modern science, be it comprehending the human genome or studying climate change, we need more talent working in science.

Let the software in

This is the cost to human opportunity that open access tackles, but a second opportunity is also waiting, foreshadowed in the declaration of the Budapest open access initiative: the chance to use the most advanced software technologies to index, transform, and search the scholarly literature to create a richer web of knowledge out of the existing scholarly canon.

Two sets of technologies await open access literature—Web 2.0 (referring to a set of tools designed to make the world wide web more collaborative, and usually taken to include weblogs, collaborative filtering, tags, and more) and the "semantic web" (a set of mark-up languages proposed by the World Wide Web Consortium). Individually and together these technologies have the potential to increase substantially the utility of information and to allow much finer grained access to information, well beyond the current state of the art of keywords and page ranking. Without access to the full text of literature they are each necessarily constrained and kept from reaching their full capability to accelerate scientific research and collaboration.

Web 2.0, a term attributed to Tim O'Reilly,⁷ represents a set of technologies that are already well embedded in popular websites and are making inroads in science. Weblogs such as In the Pipeline (<http://pipeline.corante.com>) cover the daily life of a pharmaceutical chemist, and news aggregators such as GenomeWeb (<http://genomeweb.com>) bring together news on genomics and bioinformatics. Nature Publishing Group has launched a variety of Web 2.0 sites for building communities, such as the Nature Network Boston site (<http://network.nature.com/boston>). Nature Publishing Group also runs the Connotea collaborative filtering and tagging site (<http://connotea.org>; fig 1).

Given the near exponential increase of scientific information available, this type of personalised filtering and tailored access to information represents essential future functionality. Without access to the full text, Web 2.0 is hamstrung. Readers without access to the full text are excluded from full participation in the collaborative system, and key elements of scholarly argument (such as methods and supporting data) are rarely included in abstracts.

The semantic web is a collective term for the embedding of machine readable information in the existing world wide web.⁸ It was proposed by Tim Berners-Lee, inventor of the world wide web, and developed at the World Wide Web Consortium. Designed to enrich the hyperlinking system of the web to encompass more machine readable information, the semantic web facilitates querying and inference across multiple databases (much as the web permits crawling across websites) without time consuming integration efforts. Most importantly to open access, it also holds great promise to integrate literature and databases into a unified information resource.

Just as the web required the infrastructure of the internet, the semantic web requires an infrastructure of its own. Common terms and unique names are hard to come by because they require consensus, debate, and time. There is also an enormous implicit knowledge infrastructure represented in the literature, including the connections that relate genotype to phenotype and more. Just as the full text is not freely available, neither is the knowledge infrastructure. In fact, there are subscription contracts that often forbid its extraction.⁹

The immediate value of such "public roads" for science and scholarship would be to allow computers to do tasks that should be automated: annotation of long lists of genes, protein metabolites, and chemicals with relevant information; user profile driven searches beyond the keyword; and relevance based recommendations of data and articles.¹⁰ This is why the Budapest open access initiative's call is not just for access to documents by people, but also for access that takes advantage of the rapidly changing and advancing technology.

A leading advocate of open access notes two "primary vehicles": open access journals and open access archives or repositories. The primary difference between the two is peer review: open access journals perform review while archives do not.¹¹

Open access journals typically allow the author to retain copyright, do peer review, and have different revenue models than traditional, subscription based publishers. The Public Library of Science, BioMed Central, and Hindawi are all key open access

publishers. The *BMJ* is also open access: all its research articles are freely available from publication.

Open access archives take two primary forms. The first is discipline based, the best known of which is perhaps the arXiv (<http://arxiv.org>), an online repository begun for physics. The arXiv hosts almost 400 000 papers. The second form is institutional, such as the Dspace repository at Massachusetts Institute of Technology (<https://dspace.mit.edu>), which hosts theses, articles written by its professors, and other works. Extensive research has been done into the technical barriers against providing access to the literature, resulting in both software and "how to" guides for self archiving (see www.eprints.org).

The most authoritative resource on journal policy for self archiving shows that more than 90% of journals allow some form of self archiving (see www.sherpa.ac.uk/romeo.php?all=yes). In theory then, most authors in most journals can self archive. In practice, whatever the journal states as a policy, many scientists are deterred by legal issues, which are not well understood by authors. The Wellcome Trust has commissioned a special effort to examine, clarify, and update the existing policy research.

The legal barrier to the right to archive centres around the traditional practice of copyright transfer between the authors of scholarly publications and the publishers: in return for publishers absorbing the costs of publication and peer review, the authors trade their copyrights. These copyright transfer agreements often allow authors certain rights regarding the posting of archive copies on the internet, but with a confusing forest of non-standard policies and impenetrable language. Some journals obscure their archiving policies on back pages of websites or don't publish a policy at all. Alternatives include copyright licences prepared by the Creative Commons (used by Hindawi; fig 2) and tools for authors to retain the rights of open access.

Several "addenda" to copyright transfer agreements and suggested contract language to make self archiving rights explicit have come from funders, librarians, and universities.¹²⁻¹³ These proposals provide mechanisms and tools to negotiate and modify a proposed copyright transfer agreement. To use them the author prints out and signs a copy of the one page addendum and attaches it to the agreement provided by the publisher to generate a counter offer.

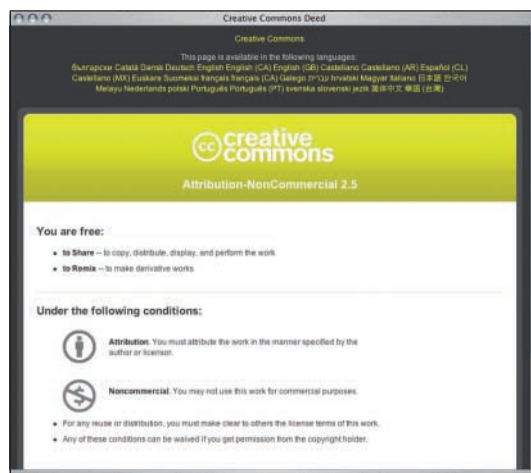


Fig 2 A copyright licence from Creative Commons

Conclusion

Authors have several resources, including open access journals, open access archives, educational materials, and legal tools, which make open access relatively easy and fully legal to achieve. Although open access has made considerable progress, and more scholarly work is publicly available than ever before, most peer reviewed articles remain closed to both human study and indexing by software. Authors, institutions, funders of research, and scholarly publishers should continue the movement towards open access so that no scholar is disadvantaged by his or her economic status and so that the full value of technological progress can be applied to the scholarly literature.

I thank Peter Suber for his comprehensive resources on open access, the eprints archive at Southampton (and Stevan Harnad's works on author self archiving), Alma Swan's economic work, the Open Society Institute open access project, the Public Library of Science, BioMed Central, Hindawi, the US National Library of Medicine's PubMed Central, the members of the Science Commons publishing working group, and all who have informed our work at Science Commons.

Sources and selection criteria: JW formerly worked on the semantic web team at the World Wide Web Consortium and founded and led to acquisition a bioinformatics company with a focus on semantics and data integration for the life sciences. He serves on the advisory board of the US National Library of Medicine's PubMed Central and is a signatory to the Berlin and Budapest declarations on open access.

Competing interests: None declared.

- 1 Budapest Open Access Initiative. www.soros.org/openaccess/read.shtml (accessed 30 Nov 2006).
- 2 Suber P. The SPARC open access newsletter. www.earlham.edu/~peters/fos (accessed 30 Nov 2006).
- 3 Morrison H. The dramatic growth of open access: implications and opportunities for resource sharing. *J Interlibrary Loan Document Delivery Electronic Reserve* 2006;16:3.
- 4 The first impact factor for *PLoS Biology*—13.9. San Francisco: Public Library of Science, 2005. www.plos.org/news/announce_pbioif.html (accessed 30 Nov 2006).
- 5 Open access journals get impressive impact factors. London: BioMed Central, 2005. www.biomedcentral.com/info/about/pr-releases?pr=20050623 (accessed 30 Nov 2006).
- 6 McVeigh M. Open access journals in the ISI citation databases: analysis of impact factors and citation patterns. Philadelphia: Thomson Scientific, 2004. www.isinet.com/media/presentrep/essayspdf/openaccesscitations2.pdf (accessed 30 Nov 2006).
- 7 O'Reilly T. What is Web 2.0: design patterns and business models for the next generation of software. www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html (accessed 30 Nov 2006).
- 8 W3C semantic web activity. Cambridge, MA: World Wide Web Consortium. www.w3.org/2001/sw (accessed 30 Nov 2006).
- 9 Elsevier license agreement. www.econ.ucsb.edu/~tedb/Journals/ucelsevier.pdf (accessed 30 Nov 2006).
- 10 Davies K. *Search and deploy*. Lancaster, PA: BioIT World, 2006. www.bio-itworld.com/issues/2006/oct/biogen-idec (accessed 30 Nov 2006).
- 11 Suber P. Open access overview. www.earlham.edu/~peters/fos/overview.htm (accessed 30 Nov 2006).
- 12 Scholar's copyright project. Cambridge, MA: Science Commons. <http://sciencecommons.org/projects/publishing/index.html> (accessed 30 Nov 2006).
- 13 Author rights: using the SPARC author addendum to secure your rights as the author of a journal article. www.arl.org/sparc/author/addendum.html (accessed 30 Nov 2006).
- 14 Willinsky J. The nine flavours of open access scholarly publishing. *J Postgrad Med* 2003;49:263-7.
- 15 Costs and business models in scientific research publishing. www.wellcome.ac.uk/doc_WTD003185.html
- 16 The best business model for scholarly journals: an economist's perspective. www.nature.com/nature/focus/accessdebate/28.html
- 17 Butler D. Open-access journal hits rocky times. *Nature* 2006;441:914. doi 10.1136/sbmj.39063.730660.F7

Startling technologies promise to transform medicine

C Donald Combs

Department of Planning and Health Professions, Eastern Virginia Medical School, PO Box 1980, Norfolk, VA 23505, USA
C Donald Combs
associate dean

Correspondence to: CD Combs
combscd@evms.edu

BMJ 2006;333:1308-11

Many writers have dared to look into the future and, far more boldly, commit their vision to print. Readers, especially those graced with hindsight, can then assess the forecast and observe the inevitable inaccuracies. I have been asked to join those ranks and cheerfully accept the challenge of forecasting some startling new technologies that will change the practice of medicine. I will focus on existing technologies that are in various stages of development and on extensions of those technologies. The five areas of technological innovation that I think will change medicine in startling ways are molecular medicine and biometrics, nanotechnology, wave technology, fabricators, and robotics and simulation (box 1). These technologies already have had individual and collective effects on some aspects of medicine and their influence is increasing.¹ Alvin Toffler was correct when he said, "Technology feeds on itself. Technology makes more technology possible."² Four brief scenarios illustrate the types of changes I envisage.

Scenarios

John, who is 10 kg overweight, walks through a device that looks like an airport x ray scanner. When he emerges, about 2 kg of fat tissue has been "fried" by a laser.³ Through the normal purging processes, the fat will be gone from his body in about three days. He

Box 1 Categories of startling technologies

- Molecular medicine and biometrics
- Nanotechnology
- Wave technology
- Fabricators
- Robotics and simulation

repeats the process every two weeks until all the extra weight is gone. No side effects are seen apart from the resizing of his wardrobe.

Jane, who has hypertension and diabetes, has a barely visible radio frequency chip implanted just below the skin on her upper arm. This chip simultaneously monitors and transmits data on her heart rate, respiratory rate, blood pressure, and concentrations of blood sugar. The data are received in two places—a remote monitoring station in her general practitioner's office, and an implanted pump that is linked to her circulatory system and to an external reservoir of drugs that can be drawn down as clinically indicated.

Robert needs knee and hip replacements. Data from magnetic resonance imaging and computed tomography scans are fed into a CAD computer program that extrapolates a customised design for the