

# IMPLICATIONS OF COPYRIGHT EVOLUTION FOR THE FUTURE OF SCHOLARLY COMMUNICATION AND GREY LITERATURE

Marcus A. Banks, Frederick L. Ehrman Medical Library, United States of America

Cees de Blaaij, Zeeuwse Bibliotheek, The Netherlands

This work is licensed under the Creative Commons Attribution-NonCommercial 3.0 Unported License. Anyone may re-use it with proper attribution, for non-commercial uses. Please visit <http://creativecommons.org/licenses/by-nc/3.0/> for further information.

## **Introduction**

Traditional practices regarding copyright are undergoing transformation. Although it is still common for scholars to give up their rights to their articles so that they will be published, this happens less frequently than it once did. Our analysis of the RoMEO database [1] shows that 75% of publishers allow authors to post their work in an online repository, whether that repository is hosted by their institution or on a personal web page. Whatever becomes of the open access movement to make all peer-reviewed journal articles immediately available online, copyright liberalization represents an enduring legacy of the open access movement.

Online repositories are a more natural home for grey literature than open access journals. Repositories can store working papers and technical reports (among other content types) just as easily as peer-reviewed articles. Crucially, repositories can also store raw data, the *grey content* that lies at the root of much scholarly discovery. Copyright liberalization has encouraged the

proliferation of such repositories; one prominent example is arXiv, which primarily serves physicists and computer scientists [2]. As scholarly discourse evolves, the preservation and promotion of grey content should command more energy than providing access to discrete grey literature.

## **I: Open Access, Self-Archiving and Institutional Repositories, and Open Data**

### *Open Access*

An open access publication is freely available to anyone with an Internet connection, and digitally archived to ensure permanent access [3]. The debate about whether to provide open access, and how, continued to evolve in 2006.

Professional societies generally support the goal of open access, which is to maximize the dissemination of scholarly knowledge. By now, the increased exposure that results from open access is empirically indisputable [4]. Despite this clear benefit, many society publishers continue to view open access publishing with ambivalence. Most societies depend on traditional subscription revenues to fund other activities, such as annual meetings. Without a comprehensive plan to replace the subscription revenues that are lost under an open access model, societies have been reluctant to embrace it. Several open access advocates have advanced proposals for how societies can surmount this challenge [5, 6].

At one time commercial publishers ridiculed proponents of open access publishing as starry-eyed idealists who did not know much about the economics of scholarly publishing [7]. Those days

are gone. In 2006 several leading commercial publishers (along with society and university publisher counterparts) began to offer a “hybrid” open access publishing option [8].

It is now possible to find open-access articles alongside traditional articles in the same electronic issue of a journal. The open access articles are available to everyone, while the traditional articles require a subscription for immediate access. The authors of each article make this decision themselves. Any fees associated with open access are absorbed by funding agencies, are waived, and are sometimes (not always) paid by the authors [9]. The hybrid model allows savvy publishers to generate several funding streams, while the traditional subscription-based model of paying for journal publication slowly contracts.

Depending upon policy developments around the globe, hybrid open access may yield to complete open access in many cases. In the United States Senate, the “Federal Public Research Act of 2006” seeks to ensure that all articles that result from research funded by the federal government, “in whole or in part,” are available for free online no later than six months after publication [10]. The bill has not passed, as of the time of this writing. It has a great deal of momentum, however, and passage in some form seems likely [11]. This is a strikingly different from the political realities in 2003, when a bill with similar aims—the “Public Access to Science Act”—was quietly buried. In the intervening years, the open access movement has matured.

The European Commission is also taking steps to endorse open access. In a wide-ranging report published in January 2006, the Commission recommends that European funding agencies “guarantee public access to publicly-funded research results shortly after publication” [12]. The

Commission provides numerous practical suggestions for how to do this. A second report, anticipated for December 2006, will expand upon this theme.

The open access movement has increased access to white literature. Nevertheless, it is important for scholars of grey literature to remain abreast of developments in this area. The Internet age is slowly eroding the traditional distinction between grey and white literature. This distinction is ultimately arbitrary, and is fading away.

In the meantime, the self-archiving and institutional repository movements bear more directly upon efforts to increase exposure to grey literature and grey content.

#### *Self-Archiving and Institutional Repositories*

The large majority of publishers allow authors to post versions of their articles on their own web site, which is known as self-archiving. Because a self-archived article is not a formally published work, it is a type of grey literature (even if the archived material is very similar to the official publication.) Although many publishers have permitted self-archiving for years, most scholars do not archive their works [13]. Institutional repositories relieve scholars of this archival responsibility, and are designed to preserve more than standard articles. For this reason, institutional repositories have great potential for increasing access to grey literature [14]. But it is not yet customary for researchers to deposit their scholarly materials (including grey materials) into institutional repositories.

## Open Data

Widespread adoption of the principles of the “open data” movement should lead to increased use of institutional repositories. The success of the open data movement will be a critical factor in shifting the focus from grey literature to grey content.

The open data movement is a corollary of the open access movement. Just as scholarly articles should receive the widest possible exposure, the data that underlies research results should also be freely available. Although the open access movement has made impressive gains, it is essentially concerned with access to the same type of information that was available in the print-only era. The open data movement is only possible in an electronic environment.

Organizations such as the Science Commons are leading efforts to increase the availability and portability of scientific data [15]. The Science Commons is an offshoot of the Creative Commons project, which allows creators of intellectual works to establish terms for the re-distribution of their work that are much more generous than standard “fair use” protections. The Science Commons seeks to instill a similar spirit about the sharing of scientific data, with the goal of “accelerating the scientific research cycle.”

Essentially, the data of interest to the Science Commons is grey content. Grey content can now be integrated into standard white literature, which is one reason why the distinction between grey and white literature is becoming moot.

The Canadian Institutes of Health Research (CIHR) released a policy in October 2006 that could greatly increase the prominence of health-related grey content in Canada. The “Draft Policy on Access to CIHR-funded Research Outputs” establishes several conditions that recipients of CIHR grants must meet. In addition to ensuring open access to peer-reviewed articles no later than six months after publication, CIHR grantees must provide public access to *research materials* and *research data* [16]. Many items qualify as research materials, including questionnaires, interview guides, and data abstraction forms. These are all examples of grey content. Research data encompasses “original data sets, data sets that are too large to be included in the peer-reviewed publication, and any other data sets supporting the research publication.” The policy specifically encourages grantees to make their research data available in an electronic form. Although they are civil servants rather than open data activists, CIHR policymakers share the same motivation for increasing the availability of grey content.

The Canadian policy is a positive development, but activism about the value of open data remains necessary on an international level. Both the United States and European Union have passed legislation that makes it more difficult to share data, not easier [17, 18]. This is designed to protect the economic interests of data aggregators. Proprietary uses of scientific data are not always inappropriate, but should be on a “value-added” basis rather than through locking away raw data that is only available via a license or other means of payment.

Some scientists have taken their own steps to increase access to data. The journal *Nucleic Acids Research* (NAR) is now fully open access, after a year as a hybrid open access journal in 2004 and many years before that as a traditional subscription-based journal [19]. Electronic versions of NAR articles often contain “supplementary materials,” which range from simple graphs to the

more sophisticated grey content of interest to open data advocates. For this paper, we evaluated whether NAR's move to complete open access produced a concomitant increase in the quantity and quality of the grey content integrated within NAR articles.

## **II: Case Study of *Nucleic Acids Research***

*Nucleic Acids Research* publishes articles about the “physical, chemical, biochemical, and biological aspects of nucleic acids and proteins” [20]. Its impact factor is in the top 10% of journals for Biochemistry & Molecular Biology, and has continued to rise since becoming fully open access. NAR articles are digitally archived in PubMed Central, beginning with the first issue published in 1974.

2002 and 2003 were the last years that NAR was published under a traditional subscription model. 2004 was a transitional year to hybrid open access, before full open access began in January 2005. NAR publishes 24 issues per year. Beginning with the first issue of 2002 until the 16<sup>th</sup> issue of 2006 (which was the most recent issue at the time of our study), we determined the percentage of “supplementary materials” that appeared in each issue. “Supplementary materials are denoted by a red flag appended to an article in PubMed Central. Not every article contains supplementary material, so the percentage equals the number of articles in each issue with supplementary material divided by the total number of articles in that issue. Yearly percentages are the average of each issue's percentage.

There has been a steady increase in the percentage of supplementary material published in NAR since 2002, with the exception of a modest decline from 2005 to 2006.

Year	Percentage of Supplementary Material
2002	11.27%
2003	16.36%
2004	22.16%
2005	31.17%
2006 (As of September)	26.42%

While the quantity of supplementary materials has increased, the caliber of these materials is of greater importance. Are they simply graphs bundled together as “supplementary,” which are not any different from what you would find in print? Or are they qualitatively different examples of “grey content” that add value to the existing article?

For eight issues published in every year since 2002 (sixteen in 2002 and eight every year since then), we sampled five articles that contained supplementary materials. Using a scale from 0 to 2, we averaged the relative greyness of the supplementary materials. 0 = no difference from what you find in a standard journal article; 1 = some difference from the content in a standard journal article, and thus some greyness; 2 = a significant difference from a standard journal article, or the highest level of grey.

Using this scale, we found a modest increase in the caliber of grey content between 2002 and 2006. This is much less pronounced than the general increase in supplementary materials.



Year	Average Greyness (From 0 to 2)
2002	0.84
2003	0.99
2004	1.075
2005	0.75
2006	1.15

### **III: Conclusion—Toward Grey Content**

The NAR case study reached a more modest conclusion than we had anticipated. While it would have been gratifying to proclaim a virtuous circle between full open access and enhanced access to grey content, the reality is that the quantity of supplementary materials increased much more substantially than the quality.

At the same time, it would have been much more sobering to report that the quality of grey content had *declined* since NAR became fully open access. NAR articles with quality grey content are examples of “datuments,” a term coined by Peter Murray-Rust and Henry S. Rzepa in 2004 [21]. A datument is a “hyperdocument” capable of “transmitting and preserving the complete content of a piece of scientific work.” By that definition, the NAR articles with top-quality grey content are certainly datuments.

The Internet is undergoing a profound transformation, from “a Web of connected documents to a Web of connected data” [22]. As this transformation unfolds, scholars of grey literature should shift their focus from managing discrete grey documents to curating diffuse grey content. The

challenge of harnessing such content is enormous, but worth the effort. Grey content is the foundation of scholarship, and we have an opportunity to make it much more accessible than ever before.

## **References**

1. SHERPA/RoMEO-Publisher copyright policies & self-archiving.  
<http://www.sherpa.ac.uk/romeo.php> [Accessed 11 November 2006].
2. arXiv.org e-Print archive. <http://arxiv.org/> [Accessed 11 November 2006].
3. Bethesda Statement on Open Access Publishing.  
<http://www.earlham.edu/%7Epeters/fos/bethesda.htm> [Accessed 11 November 2006].
4. Eysenbach G. (2006), Citation advantage of open access articles. In: PLoS Biology, 4 (5): e157, May 2006. Online at: <http://biology.plosjournals.org/perlserv/?request=index-html&issn=1545-7885> [Accessed 11 November 2006].
5. Willinsky J. (2003), Scholarly associations and the viability of open access publishing. In: Journal of Digital Information, 4 (2): Article 177, April 2003. Online at: <http://jodi.ecs.soton.ac.uk/Articles/v04/i02/Willinsky/> [Accessed 11 November 2006].
6. Prosser DC. (2004), Between a rock and a hard place: the big squeeze for small publishers. In: Learned Publishing, 17 (1): 17-22, 2004. Online at: <http://eprints.rclis.org/archive/00000945/> [Accessed 11 November 2006].
7. Hunter K. (2004), Open access: yes, no, maybe. Online at: <http://www.nature.com/nature/focus/accessdebate/3.html> [Accessed 11 November 2006].

8. Suber P. (2006), Nine questions for hybrid journal programs. Online at: <http://www.earlham.edu/~peters/fos/newsletter/09-02-06.htm#hybrid> [Accessed 11 November 2006].
9. Ibid.
10. Senate of the United States (2006), Federal Research Public Access Act of 2006. Online at: [http://cornyn.senate.gov/doc\\_archive/05-02-2006\\_COE06461\\_xml.pdf](http://cornyn.senate.gov/doc_archive/05-02-2006_COE06461_xml.pdf) [Accessed 11 November 2006].
11. ATA-Federal Research Public Access Act of 2006. <http://taxpayeraccess.org/frpaa/index.html> [Accessed 11 November 2006].
12. European Commission (2006), Study on the economic and technical evolution of the scientific publication markets in Europe. Online at: [http://ec.europa.eu/research/science-society/page\\_en.cfm?id=3185](http://ec.europa.eu/research/science-society/page_en.cfm?id=3185) [Accessed 11 November 2006].
13. Harnad S (2006), First Things First: OA Self-Archiving, Then Maybe OA Publishing. Online at: <http://openaccess.eprints.org/index.php?/archives/155-First-Things-First-OA-Self-Archiving,-Then-Maybe-OA-Publishing.html> [Accessed 11 November 2006].
14. Banks MA (2005), Towards a continuum of scholarship: the eventual collapse of the distinction between grey and non-grey literature. Proceedings GL7: Seventh International Conference on Grey Literature, 2005. Online at: <http://eprints.rclis.org/archive/00005803/> [Accessed 11 November 2006].
15. Science Commons. <http://sciencecommons.org/> [Accessed 11 November 2006].
16. Canadian Institutes of Health Research (2006), Draft Policy on Access to CIHR-funded Research Outputs. Online at: <http://www.cihr-irsc.gc.ca/e/32326.html> [Accessed 11 November 2006].

17. United States Copyright Office Summary (1998), Digital Millennium Copyright Act of 1998. Online at: <http://www.copyright.gov/legislation/dmca.pdf> [Accessed 11 November 2006].
18. European Parliament (1996), Directive 96/9/EC on the Legal Protection of Databases. Online at: <http://europa.eu.int/ISPO/infosoc/legreg/docs/969ec.html> [Accessed 11 November 2006].
19. Suber P. (2006), Nine questions for hybrid journal programs. Online at: <http://www.earlham.edu/~peters/fos/newsletter/09-02-06.htm#hybrid> [Accessed 11 November 2006].
20. Oxford Journals-Life Sciences-Nucleic Acids Research. Online at: <http://www.oxfordjournals.org/nar/about.html> [Accessed 11 November 2006].
21. Murray-Rust P, Rzepa HS. (2004), The Next Big Thing: From Hypermedia to Datuments. In: Journal of Digital Information, 5 (1): Article 248, March 2004. Online at: <http://jodi.tamu.edu/Articles/v05/i01/Murray-Rust/> [Accessed 11 November 2006].
22. Markoff J. (2006), Entrepreneurs See a Web Guided by Common Sense. In: New York Times, A1, November 12, 2006. Online at: <http://www.nytimes.com/2006/11/12/business/12web.html?hp&ex=1163394000&en=a34a6306f48166fb&ei=5094&partner=homepage> [Accessed 12 November 2006].