Towards a Continuum of Scholarship:

The Eventual Collapse of the Distinction Between Grey and non-Grey Literature

Marcus A. Banks, NYU School of Medicine, Ehrman Medical Library, United States of America

Introduction

The open access publishing movement has begun to enhance the availability of research

published in peer-reviewed journals. Valuable grey literature also deserves extensive exposure,

but in too many cases it languishes unnoticed. A common means of increasing access to grey

literature is the development of discipline-specific portals to this literature, most of which remain

fairly obscure. A more comprehensive solution would be to integrate access to grey literature

within the databases that scholars regularly consult.

These databases already identify peer-reviewed articles. The open access movement is a

political endeavor to enhance retrieval of those materials that are easy to identify but not to

obtain. With a focus on the grey literature of public health, this paper will suggest political steps

that could enhance access to materials that would often be easy to obtain if they could be

identified.

Shifting from a political to technological analysis, I will propose that institutional repositories represent an exciting possibility for both the preservation and retrieval of grey literature. The development and curation of robust institutional repositories should obviate the political struggle necessary to challenge the current hierarchy between non-grey and grey literature.

I. The Waning Primacy of Peer Review

Publishing a peer reviewed article in a prestigious journal remains the highest validation for a work of scholarship. Peer review has served scholars well for centuries. The concept of reporting and validating research findings began in 1665, with the founding of the *Philosophical Transactions of the Royal Society of London* [1]. Peer review has enhanced the rigor and relevance of many scientific breakthroughs.

In spite of its many benefits, peer review is flawed. Scholars tend to publish only statistically significant findings rather than all of their findings, which inflates the perceived importance of reported results [2]. More nefariously, business and political interests have become adept at using "research" teams to lend a patina of scientific rigor to their product development plans or policy goals [3]. These practices pervert the principle of objective scrutiny, which should be a foundation of scholarly research.

Personal computers did not exist in 1665, and for more than 300 years afterwards. The only way to disseminate research results was to print them, and this acted as a natural constraint on what

scholars could produce. Within the confines of this system, the peer-reviewed article emerged as the vector of choice for scholarship.

Personal computers are ubiquitous at academic institutions in the developed world today, and this has begun to alter conceptions of what constitutes scholarship. Digital media has substantially displaced print media, particularly in the scientific, technical, and medical (STM) domains [4]. There is no longer a pragmatic reason that the peer-reviewed journal article must stand atop the hierarchy of evidence. It is now possible to envision a seamless continuum of evidence, which is housed in well-organized and maintained repositories [5]. This continuum would still contain formal articles, which would ratify rather than validate scholarly work. Content at other points along the continuum of evidence could come to be as highly regarded as journal articles. Much of this content—such as technical reports and public policy overviews— would be what we now consider grey literature.

In this long-term vision, grey literature will become indistinguishable from non-grey literature. For the foreseeable future, this distinction will remain. Raising awareness of the value of grey literature is a political endeavor, and should be understood as such.

The open access movement, which primarily seeks to increase access to peer-reviewed articles, is explicitly political. In less than a decade, open access has blossomed from a fringe idea into a perspective that merits serious consideration [6]. Although the future of open access is unclear, it seems inevitable that the scholarly communication landscape will contain some version of

open access. Advocates for grey literature can learn valuable lessons from the political history of the open access movement.

II. Political History of the Open Access Movement

One of the most comprehensive timelines of the open access movement places its beginning in 1966, when the United States Department of Education made the Education Resources Information Center (ERIC) freely available. Open access took on political overtones in 1989, when Dr. Eddy van der Maarel resigned as Editor of the journal *Vegatatio* to protest the policies of the journal's publisher, and convinced most of the editorial board to resign as well [7]. In a statement, der Maarel explained that the intellectual labor behind the journal was free or poorly compensated, even though the journal was one of the most profitable for the Wolters Kluwer publishing company [8]. The idea that scholars should control the publishing process, rather than relying upon publishing companies, is an underpinning of the open access movement.

Throughout the 1990s, enterprising scholars in various disciplines began to publish journals that were available online at no charge. During the same period, established publishing companies invested heavily in online delivery systems. Journal prices rose sharply over this period, particularly for STM materials [9]. By the end of the decade, librarians were exasperated by the cost increases. In 1998, the Association of Research Libraries launched the Scholarly Publishing and Academic Resources Coalition (SPARC) as a means of providing cost-effective and high-quality alternatives to journals produced by major publishers [10].

In 1999, Dr. Harold Varmus proposed the development of e-Biomed, a freely accessible digital archive that would contain un-reviewed pre-prints and peer-reviewed final drafts of articles [11]. Varmus was the director of the US National Institutes of Health (NIH) at the time, and his proposal generated a great deal of criticism that the federal government was attempting to control scholarly publishing. Within a year the proposal had morphed into the politically palatable alternative of PubMed Central (PMC), a digital archive managed by the US National Library of Medicine [12]. At its inception PMC only concerned final drafts rather than pre-prints, and publishers could choose whether or not to participate.

Varmus's actions quickened the pace of calls for reform in the scholarly communication system. Although STM materials remained a flash point, scholars in all disciplines began to realize that the Internet provided a new means of developing and distributing scholarly materials. In 2000, the Association of American Universities and Association of Research Libraries released the "Principles for Emerging Systems of Scholarly Publishing." The principles state that the cost of published research should be contained, and that electronic capabilities have the potential to provide wide access to scholarship [13]. The term "open access" does not appear in the Principles, but the desire to exploit electronic capabilities is an antecedent of the formal open access movement.

The phrase "open access" gained currency in 2001, when the Open Society Institute convened the Budapest Open Archives Initiative. The Initiative states that, "The literature that should be freely available online is that which scholars give to the world with no expectation of payment."

In this spirit, it states that "open access" provides scholarly articles on the "public internet," which people can access without any "financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself." Furthermore, people are free to re-use these materials in any way they wish, provided that they acknowledge the source and "give authors control over the integrity of their work" [14].

The Budapest Initiative directed attention to the concept of open access, but did not provide a precise definition of what this means. This was the accomplishment of the Bethesda Statement on Open Access Publishing, published in 2003. The Bethesda Statement defines an open access publication as meeting two conditions: it is freely available for re-use subject to proper acknowledgement of the source, *and* is deposited in a long-term digital archive for the purpose of preservation. PMC is such a repository for the biomedical sciences [15].

Less than a week after the Bethesda Statement appeared, US Representative Martin Sabo (D-MN) introduced the "Public Access to Science Act" in the US House of Representatives. The Act sought to exclude from copyright protection any research that was "substantially funded by the Federal Government" [16]. This proposal was more politically volatile than Varmus's e-Biomed proposal. Although it did not succeed, Sabo's bill re-opened discussion about how the US federal government should facilitate access to the results of the research it funds.

In July 2004, approximately one year after Sabo introduced the Public Access to Act, the House Appropriations Committee directed NIH to *require* deposit of articles that resulted from NIH-

funded research in PMC. Grantees and publishing companies would have up to six (6) months after publication to fulfill this request. When NIH released its draft implementation policy for this plan, the "requirement" had become a "request" that grantees deposit their articles. After much consultation with various stakeholders, NIH released the official "Policy on Enhancing Public Access to Archived Publications Resulting from NIH-Funded Research" (NIH Public Access Policy) in February 2005. The policy remains a request rather than a mandate, and grantees and publishing companies have twelve (12) months to decide whether to heed the request. The policy went into effect in May 2005 [17].

Open access advocates have expressed dismay about the softening of the initial mandate. It is probably the best that can be expected in the United States, given a heavy presumption against governmental involvement in the publishing sector. To date the NIH has not received many submissions to PMC as a result of the policy [18]. This would almost certainly be different if a mandate were in effect. In October 2005 the Wellcome Trust, one of the United Kingdom's largest funders of medical research, began to mandate that its grantees deposit papers in an open access repository [19]. It will be important to compare the relative success of the US and UK approaches over time.

Many organizations lobbied for the enactment of the NIH Public Access Policy. Grey literature advocates would benefit from studying the political approach of one of these organizations in particular, the Alliance for Taxpayer Access (ATA) [20].

The ATA has put a human face on the consequences of restricted access to government-funded research. Its core membership includes people grappling with the conditions that NIH seeks to address. They ask a pointed question, which has galvanized the political discussion: If the government can spend my money to study my disease, shouldn't I be able to see the results?

For the most part, the NIH funds clinical research that results in peer-reviewed articles listed in publicly available databases. Although access barriers can preclude reading the full article, at least people know that the articles exist. The NIH Public Access Policy is an attempt to increase the proportion of this material that is available to everyone.

Public health relies significantly upon grey literature, which is difficult to identify because it is not consistently indexed [21]. The New York Academy of Medicine's Grey Literature Report is an attempt to address this challenge, but the challenge is vast [22]. As described at GL6 in 2004, the Center for Natural Language Processing at Syracuse University is developing tools for natural language searching of grey literature about public health interventions. The Robert Wood Johnson Foundation has funded this work, which is another attempt to reduce the difficulty of retrieving this material [23].

In stark contrast to the clinical research funded by NIH, there is no compelling beneficiary group to advocate for improved access to public health grey literature. Because public health benefits everyone, it appears to benefit no-one in particular. Although this perception would make it difficult, public health advocates could form an "Alliance for Public Health Promotion" in the

spirit of the ATA. Through this Alliance the representatives of leading public health

organizations would highlight the effects of a fragmented public health system on the lives and

well being of everyday Americans. With the US Congress spearheading the effort, public health

might become a greater priority for the federal government. If that occurred, comprehensive

retrieval mechanisms for public health grey literature—integrated into databases that are already

widely searched—could be developed.

Given an American political dynamic that favors limited investment in public health, this

scenario is very unrealistic at the present time. However, if public health were to become a

larger priority in the United States, increased access to grey literature would be one of many

consequences.

The goal of improving public health is not simply to facilitate access to grey literature. But it is

fair to say that the ease of access to information is one measure of how much societies value that

information. One reason for the gains of the open access movement is because many people

desire to obtain peer-reviewed clinical research without restrictions. Given the high status of

formally peer-reviewed articles, political advocacy for the complementary importance of grey

literature is essential for the near-term. Over the long-term, institutional repositories could

eliminate the need for such advocacy.

III. The Potential of Institutional Repositories to Increase Access to Grey Literature

Institutional repositories are a relatively new approach to the challenge of digital preservation of an institution's intellectual output. Disciplinary archives such as arXiv, which collects pre-prints in the fields of physics, mathematics, nonlinear sciences, computer science, and quantitative biology, are an antecedent of institutional repositories [24]. In the United States, the Library of Congress is leading a federal digital preservation effort, the National Digital Information Infrastructure and Preservation Program [25]. Self-archiving, the practice of archiving scholarly material on personal web sites, is the complete opposite of the federal approach to digital preservation [26]. Although the approaches vary widely, they all point to the urgent importance of preserving digital scholarly materials.

The self-archiving movement began in 1994, with the publication of "The Subversive Proposal" by Dr. Stevan Harnad [27]. Despite much effort by Harnad, this movement has not flourished. Researchers in many disciplines are not accustomed to taking responsibility for the preservation of their work, and have not changed their habits even after software to streamline self-archiving became widely available.

Institutional repositories shift the onus of preserving digital materials from individual faculty members to their institutions. The concept of institutional repositories became prominent in 2002, as a result of the DSpace initiative between Hewlett Packard and the Massachusetts Institute of Technology [28]. As of November 2005, over 100 institutions utilize open source DSpace software to manage their institutional repositories [29].

According to a seminal article by Clifford A. Lynch, "The development of institutional repositories [has] emerged as a new strategy that allows universities to apply serious, systematic leverage to accelerate changes taking place in scholarship and scholarly communication." Lynch defines an institutional repository as a "set of services that a university offers to the members of its community for the management and dissemination of digital materials created by the institution and its community members." Repositories can store a variety of content, including "*experimental and observational data captured by members of the institution that support their scholarly activities*" (emphasis mine) [30]. These types of data are a form of grey literature.

Faculty must be convinced of the value of contributing to institutional repositories, even if they do not have to manage them. Investigators at the University of Rochester have documented widespread faculty confusion regarding the benefits of institutional repositories [31]. This confusion resulted in smaller repository growth than anticipated for the first year after the launch of DSpace. However, there was enough activity in this period to hearten advocates of grey literature. 58% of the contents in institutional repositories during this first year were "other documents," including a healthy proportion of technical reports and working papers [32].

Interest in institutional repositories continues to increase globally. In mid 2005 repositories around the world typically contained digital versions of print-based materials, with a mixture of traditional scholarship and grey literature. Repositories in Australia and the United States conform to this pattern, but with a larger palette of content types than in most countries [33].

In the United States, institutions may begin to require that faculty members deposit their formal papers in institutional repositories. For other materials, including grey literature, deposit is likely to remain optional [34]. This is evidence of a continuing hierarchy between peer-reviewed and grey literature, but of a softer variety than before. Grey literature is not consciously excluded from institutional repositories, in the same way that it is omitted from many traditional databases.

Institutional repositories did not emerge only to facilitate access to grey literature. They are containers for capturing the complete digital output of an institution. But accessing grey literature is an important use of repositories, both now and in the future. Any organization could benefit from hosting an institutional repository. For example, non-profit research centers often produce reports that are difficult to retrieve after a short period of time. Institutional repositories could remedy this problem, and the software is available at no cost.

Content analyses of institutional repositories are critical to understanding their growth, particularly during the initial years of their development. It will soon become important to understand how researchers utilize the contents of repositories. Will people continue to gravitate to peer-reviewed articles? Or will they seek out novel content, including grey literature, because it is easier to obtain than ever before? My guess is that the latter possibility will prevail, which will lead over time to a much enriched understanding of the value of grey literature. The peer-reviewed article is an artifact of a print-based system, and not an inherently superior way to depict research results. Even if the current hierarchy between peer-reviewed materials and grey literature remains, it will be much more flat.

References

1. Royal Society: Brief History of the Society. http://www.royalsoc.ac.uk/page.asp?id=2176 [Accessed 21 November 2005].

2. Conn VS, Valentine JC, Cooper HM, and Rantz MJ. (2003), Gray literature in meta-analyses. In: Nursing Research, 52 (4) July-August 2003, pp. 256–61. – ISSN 0029-6562

3. Center for Science in the Public Interest: Integrity in Science Conference. http://www.cspinet.org/integrity/conflictedscience_conf.html [Accessed 21 November 2005].

4. Wilson B. (2003), Improving access to STM literature: the need for continuing dialogue. In: D-Lib Magazine, 9 (6) June 2003. – ISSN 1082-9873. Available online at: http://www.dlib.org/dlib/june03/06editorial.html [Accessed 21 November 2005].

5. Lynch C.A. (2003), Institutional repositories: essential infrastructure for scholarship in the digital age. In: ARL Bimonthly Report, 226, February 2003, pp. 1-7. Available online at: http://www.arl.org/newsltr/226/ir.html [Accessed 21 November 2005].

6. SPARC Open Access Newsletter: Open Access in 2003. http://www.earlham.edu/~peters/fos/newsletter/01-02-04.htm#2003 [Accessed 21 November 2005].

7. Timeline of the Open Access Movement.

    http://www.earlham.edu/~peters/fos/timeline.htm [Accessed 21 November 2005].

8. FOS Forum: Eddy van der Maarel Statement. http://lists.topica.com/lists/fos-

    forum/read/message.html?mid=902159566&sort=d&start=44 [Accessed 21 November

    2005].

9. The impact of serial costs on library collections. In: ARL Bimonthly Report, 218,

    October 2001, p. 9. Available online at: http://www.arl.org/newsltr/218/costimpact.html

    [Accessed 21 November 2005].

10. Scholarly Publishing & Academic Resources Coalition. http://www.arl.org/sparc/

    [Accesed 21 November 2005].

11. Original Proposal for E-biomed (Draft and Addendum).

    http://www.nih.gov/about/director/pubmedcentral/ebiomedarch.htm [Accessed 21

    November 2005].

12. Kling R, Fortuna J, King A. (2002), The Remarkable Transformation of E-Biomed into

    PubMed Central. Available online at: http://rkcsi.indiana.edu/archive/CSI/WP/wp01-

    03B.html [Accessed 21 November 2005].

13. Principles for Emerging Systems of Scholarly Publishing.

    http://www.arl.org/scomm/tempe.html [Accessed 21 November 2005].

14. Budapest Open Access Initiative. http://www.soros.org/openaccess/ [Accessed 21

    November 2005].

15. Bethesda Statement on Open Access Publishing.

    http://www.earlham.edu/~peters/fos/bethesda.htm [Accessed 21 November 2005].

16. Public Access to Science Act. http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=108_cong_bills&docid=f:h2613ih.txt.pdf [Accessed 21 November 2005].

17. Public Access Home Page. http://publicaccess.nih.gov/ [Accessed 21 November 2005].

18. NIH Public Access: September 2005 Submission Statistics. http://nihms.nih.gov/stats/2005-09.html [Accessed 21 November 2005].

19. Open Access a Must for Wellcome Trust Researchers. http://www.wellcome.ac.uk/doc_WTX026822.html [Accessed 21 November 2005].

20. ATA: The Alliance for Taxpayer Access. http://www.taxpayeraccess.org/ [Accessed 21 November 2005].

21. Gray BH. (1998), Sources used in health policy research and implications for information retrieval systems. In: Journal of Urban Health, 75 (4), December 1998, pp. 842-852. – ISSN 1099-3460.

22. New York Academy of Medicine Library: Grey Literature Report. http://www.nyam.org/library/greyreport.shtml [Accessed 21 November 2005].

23. CNLP: Research Projects. http://www.cnlp.org/research/project.asp?recid=34 [Accessed 21 November 2005].

24. arXiv.org e-Print Archive. http://arxiv.org/ [Accessed 21 November 2005].

25. Welcome to the NDIIPP. http://digitalpreservation.gov/ [Accessed 21 November 2005].

26. Self-Archiving FAQ. http://www.eprints.org/openaccess/self-faq/ [Accessed 21 November 2005].

27. Overture: The Subversive Proposal. http://www.arl.org/scomm/subversive/sub01.html [Accessed 21 November 2005].

28. Dspace Federation. http://dspace.org/ [Accessed 21 November 2005].

29. Dspace Instances: Dspace Wiki. http://wiki.dspace.org/DspaceInstances [Accessed 21 November 2005].

30. Lynch C.A. (2003), Institutional repositories: essential infrastructure for scholarship in the digital age. In: ARL Bimonthly Report, 226, February 2003, pp. 1-7. Available online at: http://www.arl.org/newsltr/226/ir.html [Accessed 21 November 2005].

31. Foster N.F., Gibbons, S. (2005), Understanding faculty to improve content recruitment for institutional repositories. In: D-Lib Magazine, 11 (1) January 2005. – ISSN 1082-9873. Available online at: http://www.dlib.org/dlib/january05/foster/01foster.html [Accessed 21 November 2005].

32. Ware M. (2004), Institutional repositories and scholarly publishing. In: Learned Publishing, 17 (2) April 2004 – ISSN 0953-1513.

33. van Westrienen G, Lynch CA. (2005), Academic institutional repositories: deployment status in 13 nations as of mid 2005. In: D-Lib Magazine, 11 (9) September 2005. – ISSN 1082-9873. Available online at: http://www.dlib.org/dlib/september05/westrienen/09westrienen.html [Accessed 21 November 2005].

34. Lynch CA, Lippincott JK. Institutional repository deployment in the United States as of early 2005. In: D-Lib Magazine, 11 (9) September 2005. – ISSN 1082-9873. Available

online at: http://www.dlib.org/dlib/september05/lynch/09lynch.html [Accessed 21

November 2005].