FEATURE / MANCHETTE

The Open Access Initiative, Google Scholar, and librarians: opportunity or threat?

Dan D'Agostino

Introduction

Situated on the reference desk, librarians have always acted as intermediaries between information seekers and the world of information within the library's collection and beyond. It is the unique knowledge we have gained there about how our clients search for and use information that has been the life blood of the profession. Up to this point, the rise of the digital universe has not altered this model. Library homepages, designed and organized by librarians and containing links to information sources selected and paid for by librarians, are the new intermediaries between the information seeker and the world of information. In some cases, the reference desk has even moved online. What to do then about a development that threatens to not only remove users from libraries completely, but also keep them away from library homepages? If the Open Access Initiative (OAI) succeeds and our clients no longer need the library to access material, what use will they have for librarians? This article attempts to answer that question by discussing (*i*) the latest developments in the OAI, (*ii*) in what form the initiative might succeed in fundamentally altering the information universe, (*iii*) the threat it poses to librarianship, and (*iv*) strategies necessary for the profession to thrive in a new open access world.

Open access: the background

Although the concept of open access has been around since the mid-1990s, it is still widely misunderstood as advocating the overthrow of the current scholarly communication system based on publishing in established, peer-reviewed commercial journals. In fact, the concept of open access refers only to the removal of all barriers between information seeker and information. In other words, the OAI aims to take advantage of the potential created by the development of digital publishing and the Internet to create a global online library of information that was previously collected, paid for, and held by libraries. With the price barrier removed, this open access library would be available to all users to "copy, use, distribute, transmit and display the work publicly and to make and distribute derivative works, in any digital

D. D'Agostino. Collection Development Department, University of Toronto Library, 130 St. George Street, Room 6029, Toronto, ON M5S 1A5, Canada (e-mail: d.dagostino@utoronto.ca).

medium for any responsible purpose, subject to proper attribution of authorship" [1]. Contrary to popular belief, this open access world would not mean abolishing the peer review system in favour of a kind of vanity press for academics. Nor would it necessarily entail the abolition of established academic journals for new nonprofit replacements. Finally, although generally supported by academic libraries, the OAI does not seek to rescue library budgets from the burdens placed on them by the serials crisis. Again, the only problem the OAI seeks to address is how to remove barriers to information, chiefly for the benefit of scholars, not libraries.

Advocates have described two roads leading to an open access world. The "Green" road to open access has the author self-archiving a preprint or postprint of her paper in an open access repository. The "Gold" road has the author publishing in a journal that makes articles accessible immediately upon publication. Both routes can be taken without jeopardizing the peer-review system. Authors can self-archive papers published in peer-reviewed journals — postprints if the journals allow it, preprints if they do not. New open access journals can be peer reviewed; established peer-reviewed journals can adopt an open access policy. Rather than the question of peer review, the real obstacles to the implementation of the OAI have been a lack of incentive for authors to take the Green road and the lack of a viable financial model to support publications that take the Gold road.

While peer review may not be threatened by the Gold road, the existence of subscription-supported journals certainly is. What library, already overburdened financially by the journals crisis, would pay for content that is available for free? If the subscription base of journals is significantly eroded, how will they be able to continue publishing? Although there may be many different business models that attempt to solve this problem, the one being tried most frequently at the moment is what has been called an "author pays" system. In this scenario, most notably with the many new BioMed Central open access journals, institutions are charged on behalf of their faculty for the right to publish in these journals. While removing the price barrier to accessing these articles, the "author pays" model does not address the serials crisis and has, in fact, only increased the financial burden on libraries. Not surprisingly, libraries, while generally in favour of the OAI, have been reluctant to support journals operating on this model. The viability of "author pays" open access journals remains in doubt.

Meanwhile, the development of institutional repositories, new policies adopted by major commercial publishers allowing authors to self-archive papers, and new guidelines regarding open access by grant giving agencies have made the Green road an attractive option. Many of the major commercial publishers, including Reed Elsevier, are now allowing authors to self-archive papers published in their journals in the institutional repositories many universities are now beginning to construct (the distinction among publishers at the moment is whether or not they permit self-archiving immediately upon publication or only after a set period of time). Further, grant giving agencies, such as the Wellcome Trust, are now requiring grant recipients to publish open access versions of their articles. Last winter, in what might have proved the greatest boost yet for the OAI, the National Institutes of Health (NIH) considered a similar proposal. But after intense lobbying by commercial and society publishers, the NIH backed down and in the end asked authors to voluntarily submit their articles for open access inclusion on PubMed Central within 12 months of publication.

A way forward

Far from being the end of the OAI, the new NIH policy may point to a method for making established journals open access in a way that maintains current subscription bases. Since most scientific articles are read within the first year of publication [2], journals can protect their subscription base by releasing articles to open access repositories only after an embargo period of at least 6 months. I think a fair assumption is that any library that is willing to pay to acquire content within the first year of publication would want to pay for content within the first 6 months of publication; which is to say that an embargo period of 6 months would not lead to a significant decrease in the number of a journal's institutional subscribers. (There is one important caveat: this may not be the case for journal packages. Universities may balk at paying for titles whose content they are willing to wait 12 months for, and this might provide libraries with the necessary leverage to make package deals obsolete. This in turn would lead to substantial revenue loss for the big commercial publishers who have made large profits by forcing libraries to subscribe to low impact journals that traditionally cost very little to produce.) Although this solution does nothing to address the need for open access immediately upon publication and may do little to alleviate the journals crisis and the damage being done to library budgets, it has the virtue of being achievable. However, the fact that a way forward does exist does not mean that academics will embrace the idea of open access when it comes to putting their own papers in repositories. Two recent developments at Google may provide researchers with the incentive they need to begin to self-archive in a serious way.

In 2004, Google made two surprise announcements. The first was Google Print, a project to scan and mount full-text versions of the collections of several major academic libraries and then make the material available for free on the Web. This development has been much commented on by librarians who feel it may take users out of libraries forever and by those who dismiss it outright as unachievable. As it relates to the current state of the OAI, Google Print's most positive effect is to provide momentum to the idea that the dream of a barrier-free online library of the world's information is both desirable and achievable. It is, rather, the second initiative, a search engine designed to search the academic literature called Google Scholar that has been less commented on by librarians but should be of much greater concern.

Scientists publish papers for many reasons, but none more important than to provide visibility to their work. Visibility is ensured by publishing quickly in high impact journals; visibility is measured in how often their work is cited. The genius of Google Scholar is not only the comprehensiveness of the publications searched, but the fact that it functions like Science Citation Index and Scopus as a citation index. That is, it provides links to papers that cite the paper the user is viewing. As a citation index, it is able to provide a measure of how important a paper is by determining how often it has been cited. Although the logic behind measuring a paper's importance in its field by how often it is cited may be dubious, the concept has been embraced by faculty and university administrators over the past several years. Putting this functionality in a Google product, a search engine already widely favoured by faculty, should ensure the popularity of Google Scholar with researchers. A high use of Google Scholar among researchers will in turn promote the adoption of open access among this same population in two ways. First, by showing that those who publish in open access journals or who self-archive are more likely to do better in the citation counting competition, since early studies are showing that open access articles are cited more than those in subscription only journals [3]. Second, articles whose full text is available through Google Scholar are more likely to be read by the many faculty, university administrators, administrators of grant giving agencies, journalists, and members of the public who will be using Google Scholar than articles that can only be accessed through a subscription. It is even possible that the increase in visibility gained by having an open access paper on Google Scholar will put pressure on established journals to significantly reduce embargo periods.

To sum up, the prospects for the creation of an open access world continue to look positive. But what initially emerges will not be largely open access to information immediately upon publication, and this means that libraries will continue to provide their clients with access to current subscription-based publications. Google Scholar, if it succeeds in becoming the first search tool of choice among researchers, will provide the necessary incentive to authors to mount open-access versions of their papers in institutional repositories, as soon after publication as their journals allow. The question then is what impact will this less ambitious open access world have on libraries and librarians?

A role for librarians

Although scientists in most disciplines are reading more articles than ever before and relying on library-provided access to do so, a surprising number remain unaware of the role libraries play in providing these journals [4]. As Google Scholar allows institutions to provide their users with links from the search engine to the full text they pay for, the notion that their institution's librarians are not involved in their

research will only increase. Therefore, is there a role we can perform for our faculty other than simply providing content to link to Google Scholar?

Two developments in scholarly communication that are occurring parallel to the OAI may have the paradoxical effect of making us more relevant than ever, just at the time when our clientele thinks they are leaving us behind. The first is that scientific publishing is becoming more data intensive. Data, in the form of large databases like the Human Genome Project, are being increasingly shared and utilized by researchers. Indeed, some, such as Vitek Tracz, the head of the Current Science group, believe that the sharing of data will become so important that many journals may be replaced in the future by large databases that organize data in meaningful ways and contain text commentaries:

> Above all, however, I believe it means highly specialized, editorially intensive databases — databases that take data, usually publicly available data, and put it together in a structure that makes it more useful and understandable by organizing it and adding commentaries and analysis. So, where today you have thousands of journals sold on subscriptions, in 10 to 20 years there will be thousands of editorially intensive databases also sold on subscriptions, many of them probably sold by existing science publishers [5].

Whether or not many journals will eventually disappear is unclear, but the key concept here is that scientists will need assistance in creating, managing, and searching these databases. Who better to help than librarians, who have always had unique knowledge of how scientists search for and use information?

The second development is that scientists continue to need to read more to keep up with their discipline but do not have more time to read than in the past [6]. Again this implies that there will be a need for tools (beyond those that currently exist) to help scientists search the literature as efficiently and effectively as possible. These tools will have to be developed at the subject level to be helpful, that is, responsive to the specific subject needs of researchers and beyond the capabilities of Google Scholar, since citation counts do nothing to tell you how important a paper is within the first year of publication. Evidence that such a need exists is that researchers themselves are beginning to create tools and make them available for free. One such tool is CiteULike (see http://www.citeulike.org/), a free online service where researchers can organize the citations they are reading as bookmarks, tag them with keywords, and share that information with other researchers using the service. Knowing what other researchers in your field are reading helps to direct you to must-read articles in your discipline. CiteULike goes one step further than simply providing a bookmark organizing service: they now allow researchers to create their own per-

sonal library of articles in PDF format mounted on the CiteULike server (to comply with copyright, users must first download the article to their personal computer and only then to the CiteULike server). The fact that researchers themselves are beginning to develop these tools should alert librarians everywhere that these needs exist and that if we do not take the initiative in addressing them, scholars will for themselves.

The key to survival: getting to know our users

The creation of tools that are suited to the needs of specific users for sorting through databases and literature is, in fact, what librarians have always done. Before the Internet, we did this at the reference desk, for example, by guiding users to indexes to the journal literature. In the first 10 years of the Internet, we have adapted the same approach by guiding our users to tools and literature we have collected on the library's homepage. In a largely open access, Google-dominated world, the only difference will be that we are creating these tools ourselves for literature that the library has not always selected.

The key to creating this new role for ourselves is actually knowing first hand what information researchers are searching and what tools they are searching with. In the past, it was our position on the reference desk that gave us this knowledge. If researchers are by and large avoiding libraries, we can only now regain this knowledge by going into research offices. We will need to know how their information-seeking behaviour is changing, while it is changing. We will need to sell them on our ability to create resources that help them do their research more efficiently. To remain in the library is the one way to guarantee that our role will shrink from that of the unique intermediary between information seeker and information that we have always been, to that of essentially clerical staff maintaining the financial end of a static collection of journal subscriptions.

References

1. Bethesda statement on open access publishing. 2003 Jun. Available from http://www.earlham.edu/~peters/fos/bethesda.htm.
2. King DW, Tenopir C, Montgomery CH, Aerni SE. Patterns of journal use by faculty at three diverse universities. *D-Lib Magazine* [serial on the Internet]. 2003 Oct;9(10). Available from http://www.dlib.org/dlib/october03/king/10king.html.
3. The effect of open access and downloads ('hits') on citation impact: a bibliography of studies. Highfield, UK: OpCit Project. Available from http://opcit.eprints.org/oacitation-biblio.html.
4. King et al., op. cit.
5. Poynder R. Essential for Science: Interview with Vitek Tracz. *Information Today.* 2005 Jan;22(1):1, 6.
6. King et al., op. cit.