The International Journal of Digital Curation
Volume 8, Issue 1 | 2013

The Critical Role of Institutional Services in Open Access Advocacy

Tomasz Neugebauer and Annie Murray,
Concordia University

Abstract

This paper examines the development of the Open Access movement in scholarly communication, with particular attention to some of the rhetorical strategies and policy mechanisms used to promote it to scholars and scientists. Despite the majority of journal publishers' acceptance of author self-archiving practices, and the minimal time commitment required by authors to successfully self-archive their work in disciplinary or institutional repositories, the majority of authors still by and large avoid participation. The paper reviews the strategies and arguments used for increasing author participation in open access, including the role of open access mandates. We recommend a service-oriented approach towards increasing participation in open access, rather than rhetoric that speculates on the benefits that open access will have on text/data mining innovation. In advocating for open access participation, we recommend focusing on its most universal and tangible purpose: increasing public open (gratis) access to the published results of publicly funded research. Researchers require strong institutional support to understand the copyright climate of open access self-archiving, user-friendly interfaces and useful metrics, such as repository usage statistics. We recommend that mandates and well-crafted and responsive author support services at universities will ultimately be required to ensure the growth of open access. We describe the mediated deposit service that was developed to support author self-archiving in Spectrum: Concordia University Research Repository. By comparing the number of deposits of non-thesis materials (e.g. articles and conference presentations) that were accomplished through the staff-mediated deposit service to the number of deposits that were author-initiated, we demonstrate the relative significance of this service to the growth of the repository.

International Journal of Digital Curation (2013), *8*(1), 84–106. http://dx.doi.org/10.2218/ijdc.v8i1.238

Introduction

The movement to bring about Open Access to scholarly and scientific literature has many adherents in the science, technical and medical disciplines, and slowly growing numbers of adherents in other disciplines. Members of the library community have long espoused the societal value of openly available scholarship, and have been active in the call for the re-imagining of scholarly communication since university library acquisitions budgets have not been able to sustain the continuing increase in both the number and cost of scholarly journals.

The relationship between open access to research information and the quality of research information, and ultimately science itself, has emerged as an ongoing theme in discussions about the usefulness and promise of open access. Does increased access to research have any bearing on the quality of the research that is subsequently produced? Is open access merely a rethinking of dissemination or can it drive or hasten improvements and innovation in science? There is no doubt that increased worldwide access to science is a good thing, but does increased access increase the quality of the science being published?

Increasingly, funding agencies have been adopting policies that oblige scholars and scientists to make their research openly available in open access repositories. Likewise, a growing number of universities around the world mandate their researchers to archive copies of their work in an institutional repository.

The continuing success of the open access repository arXiv and the promise of OA journal publishing ventures, such as the titles published by the Public Library of Science (PloS), indicate the desirability and value of unfettered communication of research results and the willingness on the part of scientists, institutions and publishers to explore new models of scholarly communication. Open Access has made significant inroads in a gradual rethinking and reforming of models of scholarly dissemination. However, the movement is characterized by competing visions of how open access to scientific literature ought to be brought about. Moreover, surveys of faculty members continue to reveal ongoing concerns about open access. In order to persuade researchers to adopt open access in their research practices, different rhetorical strategies have been employed by institutions, the scientific community and advocates of open access.

In the absence of mandates from funders or universities, advocates of open access use different lines of reasoning to persuade scholars to make their work openly available on the web. Librarians and administrators of repositories appeal to the desire of scholars to have their work highly visible and thus more likely to be cited. Open access advocates also appeal to the desire for equal access to information, rightly arguing that people from all institutions and countries, regardless of the number of periodical subscriptions held at their college or university library, should have equal access to published research. Thus advocates often appeal to the scientist or scholar's sense of personal visibility and career progress, while simultaneously appealing to the more altruistic aim of sharing knowledge and giving back to others who could not otherwise afford access.

Another theme evoked, sometimes problematically, to promote greater adoption of open access is an appeal to scholars to move science or knowledge itself along further by creating an ever larger, fully searchable, open body of scientific research on the web. It is argued that this will increase readership, citation and generally help speed up innovation and the development of new ideas and discoveries. We discuss the problematic rhetoric of open access as a means to the end of innovation through text/data mining and the development of semantic web services later in this paper. Certainly, science cannot flourish in a vacuum; expensive science journals only available to the richest of institutions contribute to an expensive information silo. The value of openness itself has emerged as a desirable goal for scientific publications, with the Public Library of Science playing a leadership role in offering web-based, high quality, peer-reviewed scientific information at no cost to the reader or user. However, promoting open access as a necessary step towards innovation through text/data mining and the development of semantic web technologies is not universally effective and convincing. In advocating for open access, we recommend focusing on its most universal and tangible purpose: increasing public (gratis) open access to the published results of publicly funded research.

Clifford Lynch's 2003 description of the institutional repository has informed our approach:

> "In my view, a university-based institutional repository is 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."(Lynch, 2003).

Researchers need strong institutional support to understand the copyright climate of open access self-archiving, user-friendly interfaces and metrics, such as repository usage statistics. We recommend that mandates, together with well-crafted and responsive author support services at universities, will ultimately be required to ensure the growth of open access. In this paper we describe the mediated deposit service that was developed to support author self-archiving in Spectrum: Concordia University Research Repository. By comparing the number deposits of non-thesis materials (e.g. articles, conference presentations, etc.) that were accomplished through the mediated deposit service to the number of deposits that were author-initiated, we demonstrate the relative significance of this service to the growth of the repository.

Open Access Advocacy

Access, Quality and Peer Review

Although there is much discussion and debate regarding the peer review process (Sieber, 2006; Harnad, 2000) as well as empirical studies about its effectiveness (van Rooyen, Godlee, Evans, Smith & Black, 1998; Seglen, 1997), peer review continues to be perceived and presented as the key or central mechanism that ensures the quality of academic publishing (Brinn, Jones & Pendlebury, 2000; Ware, 2008; Harnad, 2000). The IEEE Computer Society summarizes this position with the following statement:

> "The IEEE Computer Society has a rigorous peer review process in place to ensure the high quality of its technical material" (IEEE, 2011).

Although there are many factors that contribute to a research article's quality (e.g. accuracy, validity, and reliability of measures and references), in the publishing phase of the research process, peer review is the mechanism that ensures quality. As defined by Mario Bunge, a "mechanism is a set of processes in a system such that they bring about or prevent some change – either the emergence of a property or another process – in the system as a whole" (Bunge, 2003).

It is important to point out that although the majority of the published results provide evidence that open access articles are cited more (Brody, Harnad & Carr, 2006; Gentil-Beccot, Mele & Brooks, 2010; Lawrence, 2001), and citation counts are conventionally used to measure quality, "OA itself will not make an unusable (hence uncitable) paper more used and cited" (Gargouri et al., 2010). Peer review and open access are both mechanisms of the publishing phase of the research process. Certainly, peer review is a mechanism that ensures quality. Open access is a mechanism for the distribution of research, not a mechanism that ensures quality. If it were true that open access increases research quality, then only a weaker claim could be true regarding peer review: it is one of the mechanisms that ensure the quality of academic publishing.

Open Access Mandates and Repositories

The development of repositories has seen the establishment and growth of subject-based and institutionally-based repositories. Some examples of subject-based repositories include the arXiv e-print server in high-energy physics[1], E-LIS[2] in the field of Library and Information Studies, and PubMed Central[3] in biomedical research. Examples of institutionally-based repositories include, for example, University of Southampton's repository[4], or the repositories of any number of universities around the world, among them MIT[5] or University of the Arts London[6].

Despite the existence of many repositories and the general acceptance of author self-archiving by publishers, it has been estimated that only 19.4% of refereed journal articles are available in open access repositories (Bjork, Roosr & Lauri, 2008) and it is not uncommon for a repository to contain very few items. Harnad and McGovern suggest that:

> "Institutional repositories within higher education can only be counted as successes if they are ingesting a significantly higher percentage of their institution's scholarly output, approaching 100%, rather than languishing at the global baseline [15-20%]" (Harnad & McGovern, 2009).

[1] arXiv e-Print Archive: http://arxiv.org/

[2] E-prints in Library and Information Science (E-LIS): http://eprints.rclis.org/

[3] PubMed Central: http://www.ncbi.nlm.nih.gov/pmc/

[4] ePrints Soton: http://eprints.soton.ac.uk/

[5] Dspace@MIT: http://dspace.mit.edu/

[6] UAL Research Online: http://ualresearchonline.arts.ac.uk/

An open access mandate or policy requires authors to deposit the results of their research in an open access venue: an institutional repository in the case of an institutional mandate, or in a subject or institutional repository in the case of a funder mandate. ROARMAP[7] is a registry of open access policies and mandates; at this time, there are over 150 institutional mandates and over 50 funder mandates in place.

As Kennan helpfully summarizes:

> "In analyzing the relationships and entanglements that exist between authors, universities, publishers, and other actors we see how these reinforce the current publishing paradigm unless a clear effort is made to achieve and support change. It takes a new actor, such as the mandate or deposit policy, to encourage some universities and authors to look beyond their existing frames." (Kennan, 2011).

Arthur Sale (2006) has shown that the key factor in having faculty members deposit in their institution's repository is a mandate, and for many open access advocates, a mandate is seen as the most important mechanism to bring about greater access to published research. Yet there is not universal agreement that a mandate will bring about such change. In Peter Linde's reflection on the open access policy of his institution, Sweden's Blekinge Institute of Technology (BTH), he concluded that:

> "BTH's OA-policy has neither meant a dramatic increase in published full-text articles in our publication database, nor in OA-journals, in spite of the support for parallel publishing the library has offered since a year back" (Linde, 2010).

While Kennan reasons that "without a mandate the OA message is ambiguous, it does not appear as if the university has unconditional support for OA or its IR" (Kennan, 2011), it is interesting to consider Linde's observation on the need to "anchor and supplement our policy document with a descriptive and constructive guide, telling individual researchers at BTH what they can do to fulfil [*sic*] different stipulations and wishes regarding OA" (Linde, 2010).

We can see then that institutional commitment to open access and the growth of repositories may be signaled by the adoption of an open access policy, mandate or other legal mechanism, but the deeper institutional commitment is more deeply revealed in the policy's subsequent implementation and support through the services and infrastructure that empower authors to become active participants in open access. Nancy McGovern noted that:

> "Mandates can be difficult to sustain in the absence of dedicated funding from the institution for which it is implemented. The institution effectively commits to the idea rather than the reality of having the repository" (Harnad & McGovern, 2009).

Without the development of this socio-technical infrastructure, the dynamics of scholarly communication remain largely undisturbed for a particular subject domain

[7] Registry of Open Access Repositories Mandatory Archiving Policies (ROARMAP): http://roarmap.eprints.org/

or for the authors at a specific institution. McGovern cites faculty resistance and "unfunded" mandates as factors that can negatively impact the success of an institutional repository (Harnad & McGovern, 2009).

There is some debate about the role of a subject repository versus an institutional repository with respect to mandating a place of deposit. It may be more meaningful for a scholar to have his or her work stored alongside the work of the colleagues in his or her field rather than being grouped with the colleagues at his or her home institution. Steven Harnad argues that institutional mandates ought to specify the institutional repository as the place of deposit, since central/subject repositories can harvest from these (Harnad, 2008). On the surface, this sounds like a practical solution to a complex problem. However, if we consider scholarly communication as a system, then the introduction of a subject repository is what mimics more closely the communications role that the scholarly journal has been serving. The relations between scholars and their audience form an organization mediated through the scholarly journal, but the institutional repository as a place of deposit has a radically different organizational structure and demands that the researcher considers a new audience that is more like a crowd than the discipline-specific organization of scholars that access a particular journal. We see no need to mandate the type of open access repository in which researchers ought to deposit their work. If it is more in keeping with the disciplinary practices for a researcher to place work in a subject-based repository, then forcing him or her to deposit in the local institutional repository will be a very tough sell. If the goal of a mandate is to make publications openly available, a usable and well-staffed institutional repository should be provided, but it should not be mandated as the sole place of deposit. The location of the open access repository, like the decision about the specific journal in which to publish, should be the researcher's choice.

Let's consider the scholarly communication as a complex object composed of individual researchers and publications. Mario Bunge explains that "a property of a complex object is said to be emergent if neither of the constituents or precursors of the object possess it" (Bunge, 2003). Individual scholarly journals, then, are complex objects composed of individual articles and, as such, they have emergent properties including but not limited to the journal's impact factor, editorial policies or research quality.

Bunge distinguishes between crowds and organizations (systems) in that the latter have structure, "the collection of relations among its components" (Bunge, 2003). Individual journals have structure as business firms and scholarly associations. A bonding relation, such as a social relation, "makes a difference to the relata", whereas a non-bonding relation (such as spatial 'to the left of') does not. Scholarly communication consists of many organizations with bonding relations and emergent properties.

Although the financial dependence of scholars on their institutions is a strong bond, an institution forms a small system compared to the large social systems of subject-specific research communities that span across institutions, countries and continents. In terms of scholarly communication, these larger systems are perceived as more significant to the researchers than the small institutional systems formed by the financial bonds of academic salaries and institutional colleagues.

Mandating the place of deposit to be solely the institutional repository rather than a subject repository is contrary to the current structure of scholarly communication. Researchers publish with a particular audience in mind, the audience forming an emergent property of the place of publication. As Foster and Gibbons argue:

> "When it comes to research, a faculty member's strongest ties are usually with a small circle of colleagues from around the world who share an interest in the same field of research" (Foster & Gibbons, 2005).

The institutional repository lacks the particular audience of a publisher, such as a scholarly journal, association or busy subject repository. A researcher requires a leap of faith that the benefits of deposit will be worth the effort because it is more difficult to predict or imagine in the case of an institutional repository, rather than subject-based repository or research journal (which is similar to a subject-based repository), what new social bonds will result.

It is important to recall, as Jingfeng Xia does, that:

> "Subject repositories are founded or operated by scholars. With their understanding of the need of scholars as individuals and the scholarly society as a whole, these OA advocates have been able to build a sustainable program that fits into the established communication system and attracts passionate people to make active contributions" (Xia, 2011).

This could go a long way to explain the reluctance of scholars to embrace depositing in institutional repositories, since there may be no compelling scholarly reason to do so. It may feel as though a deposit into an institutional repository could be an off behavior for a scholar who is tied into a specific disciplinary culture with its own existing norms.

This is not to suggest an institutional repository does not fulfill a critical function at a college or university. Rather, an institutional repository should work in concert with subject repositories, offering a scholar or scientist a choice in the place of deposit for their work. Institutional repositories can serve as key tools for the preservation and dissemination of research results for disciplines with open access policies or mandates in place, and especially for disciplines without subject repositories.

Is the new, unknown institutional repository audience a threat to the current organized social structure of scholarly communication? Indeed it may be an opportunity, since there is strong evidence that open access articles are cited more. Presumably the audience of institutional repositories overlaps considerably with the audience of subject-based repositories and the journal in which the article was originally published. The key factor here may simply be the critical role of Google and Google Scholar as the new mediator in indexing and delivering scholarly content, regardless of the type of repository from which the article originated.

Librarians and open access advocates fully appreciate the difficulties and inequalities caused by journal pricing, and thus promote institutional repositories as a centralized place for collecting and disseminating a university's research. However,

do scholars themselves see the value of contributing their work to an institutional repository? In 2003, Jean-Claude Guédon predicted that:

> "Open access archives will succeed only if they can satisfy the scientists' needs in terms of their careers. This means incorporating certification or branding tools that are at least as authoritative as those presently provided by existing journals" (Guédon, 2003).

At this point in time, it is doubtful that an institutional repository can offer this kind of immediate relevance to a scientist's career, even if a mandate is in place. This might explain why faculty participation in institutional repositories remains low: despite the good intentions of librarians and many in the open access movement, scientists do not yet feel sufficiently motivated to contribute copies of their work to an additional scholarly communications node since they perhaps perceive they have done enough to disseminate their work in the journal or disciplinary repository that makes sense to them. It may be that faculty members are not sufficiently convinced of the real utility of institutionally-based repositories, despite convincing evidence that doing so will increase visibility and citations.

Why don't scientists sufficiently exploit the wide dissemination offered by institutional repositories? Perhaps open access advocates and managers of institutional repositories need to adjust rhetorical strategies currently employed to entice faculty members to make their work openly available by appealing to the huge new institutional repository audience. Peter Suber suggests that "one of the best-kept secrets of scholarly communication today is that deposit in an OA repository is comparable with publication in a TA [toll access] journal" (Suber, 2010). Suber explains that many authors assume that publishing in a TA journal and then depositing in an open access repository are incompatible. For instance, they may not think that deposit is allowed by publishers. However, given the increase in usage and citation that occurs when there is a freely available copy of an article in a repository, we can easily see how compatible the two practices really are. But why is this compatibility not better understood or exploited? Suber suggests: "If we could enlighten researchers and their institutions on this one point, we'd remove one of the largest single barriers to the spread of OA" (Suber, 2010).

In order to show researchers the benefits of deposit in an institutional repository, we believe that the statistics modules offered by most repository software are a crucial infrastructure that demonstrates to authors that their work is finding new audiences. Use and download statistics available from institutional repository software offer a vital service to faculty: proof that their work is being accessed and used. A new audience and a new metric are emerging, but how long will it take to make a tangible difference in the careers of scientists? It would seem that appeals to equality of access and altruism have only enjoyed minimal success, hence the only marginal success of institutional repositories. As many repository managers have learned, ideals cannot populate a repository. Rather, the repository has to make sense in the context of the researcher's career. How can the two systems and audiences intersect more holistically in a way that will demonstrate to researchers that there are compelling benefits associated with depositing their work in an institutional repository?

The Budapest Open Access Initiative's recently introduced *BOAI 10* document makes recommendations on how to "set the default to open access" in the next ten years (i.e. 2012-2022). Their first recommendation is that "every institution of higher education should have a policy assuring that peer-reviewed versions of all future scholarly articles by faculty members are deposited in the institution's designated repository" (BOAI, 2012) further solidifying the prevailing belief that mandates are the main way to bring about greater access to research. Moreover, the BOAI promotes the role of the institutional repository as a key campus tool not only for access to and preservation of research materials, but they also suggest that "universities with institutional repositories should require deposit in the repository for all research articles to be considered for promotion, tenure or other forms of internal assessment and review" (BOAI, 2012).

Thus we can see the ongoing tension in open access strategies due to differences in ideas about the ideal place of deposit and the role of the institutional repository for a particular university and for scholarly communication in general. Clifford Lynch's clear-sighted positioning of institutional repositories as "essential infrastructure" for scholarship, and BOAI's urging for the necessity of institutional repositories to "set the default to OA" are certainly above reproach. Yet, scholars by and large have been reluctant or unable to seize upon institutional repositories as a means of giving their research barrier-free accessibility and high visibility. Do the reasons for lack of uptake in self-archiving suggest an inherent limitation in the true service capacity and symbolic value of institutional repositories? Beyond mandate, what will truly initiate change in well-entrenched systems of scholarly communication?

Measuring Attitudes Towards Open Access

Academics and research funders have expressed a commitment to open access by signing various petitions, declarations and statements in support of open access over the course of the last two decades (Open Access Directory, 2011). One of the first petitions by the Public Library of Science in 2001 suggested a boycott of all journals that did not offer OA after six months. Although it was signed by 34,000 biomedical researchers, the boycott did not materialize. The petition to the European Commission for guaranteed public access to publicly-funded research results[8] has been signed by over 28,000 people since January 17th, 2007. Some University faculties, such as Harvard University Faculty of Arts and Science, have passed resolutions requiring its members to self-archive their peer-reviewed research. There is evidence that existing mandates have been successful in filling institutional repositories (Sale, 2006).

Studies reveal that a majority of academics (81%) "would willingly comply" with an institutional mandate to deposit their research (Swan & Brown, 2005). The important point to remember, however, is that using surveys to measure what is in people's minds is an imperfect art. For one, the way in which the questions are formulated can have a significant impact on the results. As Bunge summarizes:

> "It is true that interviews and opinion polls constitute windows on people's opinions and intentions. But they yield restricted and ambiguous indicators, because the pollster may not ask pertinent and important questions" (Bunge, 2004).

[8] Petition for guaranteed public access to publicly-funded research results: http://www.ec-petition.eu/

Asking if someone would willingly comply with a mandate is something entirely different from asking them if they would vote to adopt a mandate, or why they choose not to self-archive. Although the number of faculty imposed mandates is growing, the majority of university faculties have yet to pass them. Further, once a mandate is established, the task of ensuring faculty participation or compliance is another matter entirely – one that is fraught with bureaucratic nuance and which touches upon questions of resource scarcity facing most repositories and universities.

Researchers have the option to support open access by self-archiving but, without a mandate, only a small percentage choose to do so (Harnad et al., 2004). Open access advocates attempt to manage the social change through mandates, but there is a clear need for surveys that help to explain the current reality of reluctance to self-archive. Some recent studies shed some light (Kim, 2010; Morris & Thorn, 2009; Fry et al., 2009). Morris and Thorn find that a significant number of researchers do not voluntarily self-archive because they are worried about the negative implications (for themselves as well as the journals in which they publish) of making available multiple versions of their articles online, e.g.

- 'I do not want non-peer-reviewed versions of my articles to be in circulation' (31.02%);
- 'I do not want non-copy-edited versions of my articles to be in circulation' (18.67%);
- 'I do not want multiple versions of my articles to be in circulation' (26.08%);
- 'I do not want to damage the journals in which I publish' (21.14%). (Morris & Thorn, 2009).

Kim limited his study to 17 non-mandated deposit doctorate universities and found two factors that impede self-archiving: concerns about copyright and additional time and effort required to self-archive (Kim, 2010). Fry et al., (2009) confirm that: "Despite the SHERPA/RoMEO service, there is often confusion amongst authors regarding their rights and responsibilities in relation to copyright agreements with publishers, and again this level of understanding appears to vary community by community." These faculty concerns are persistent, pervasive and legitimate.

The Morris and Thorn survey confirmed that a significant percentage of those that do self-archive, do so because it is mandated by their funder, institution or department (Morris & Thorn, 2009). In another faculty survey, researchers from the Ithaka[9] organization concluded:

> "Despite several years of sustained efforts by publishers, scholarly societies, libraries, faculty members and others to reform various aspects of the scholarly communications system, a fundamentally conservative set of faculty attitudes continues to impede systematic change" (Schonfeld & Housewright, 2010).

They note also that:

[9] Ithaka: http://www.ithaka.org

> "...although faculty attitudes are only one component of policy making for scholarly communications, they may help to explain why policy makers have in some cases turned towards incentives or deposit mandates. Without this kind of interest and investment from university leadership, changes to the scholarly publishing system are likely to happen slowly, if at all."

Mario Bunge summarizes the idea that opinion polls are insufficient to explain systemic change:

> "Obviously, different people are bound to feel somewhat differently about the prevailing social order. In particular, some will dislike it because it hurts their interests, others because they find it stale, still others because they find it unfair, and so on. However, knowledge of such subjective factors is insufficient to explain revolutions. Revolutions succeed not just when many people object to the establishment: they succeed when the revolutionary elite – which always includes some members of the establishment – manages to mobilize the part of the population that is deeply dissatisfied with the prevailing regime, while the latter is not prepared to counter the rebellion." (Bunge, 2004).

Although opinion polls and surveys are necessary, "to keep the democratic mechanism running" (Bunge, 2004) a foundational argument for open access is that knowledge is a public good (Suber, 2011). Since the State has the mandate of enhancing everyone's quality of life by managing and delivering public goods (Bunge, 1998), it follows that institutions that yield political power within academia have the responsibility to ensure that research output is managed wisely through the use of policy, and this can include self-archiving mandates.

If open access is a systemic, revolutionary change in scholarly communication, the problem seems to be the lack of visibility and mobilization of the population that is deeply dissatisfied with the prevailing regime. Although the revolutionary elite is quite visible and mobilized, the public protest against the lack of access to knowledge in scholarly journals seems imperceptible. Indeed, the information age has brought with it an explosion of content to the public which was previously inaccessible – and it is growing exponentially.

OA Advantage

Aside from subjective measures of the popularity and desirability of open access policies, there are also the more objective measures of the open access providing an increase in research impact.

There have been a number of studies published on the so-called "OA advantage" over the course of the last decade; the Open Citation Project maintains an up-to-date annotated bibliography of studies on the relationship of open access downloads and citation impact (Open Citation Project, 2012). In addition, annotated bibliographies of studies on the open access impact advantage have been published recently (Wagner,

2010; Swan, 2010). Although the majority of the published results provide evidence that open access articles are cited more (Brody, Harnad & Carr, 2006; Gentil-Beccot, Mele & Brooks, 2010; Lawrence, 2001), the literature also includes empirical studies where the OA citation advantage is not replicated (Davis, 2010; Davis, 2009; Davis, 2011; Wagner, 2010).

There is also a discussion about the nature of the advantage, such as whether or not it is caused by a self-selection bias (Gargouri et al., 2010). Are authors are self-archiving only their best, more citable work, and thus causing the studies to show a citation advantage for the open access articles? Or, is it true instead that making an article available on an open access platform actually causes the article to become more citable by the nature of the fact that it is now available to more scholars?

Although these studies are interesting challenges in bibliometric statistical analysis, they remain unconvincing as a motivator for researchers to undertake the effort required to make more of their research openly accessible. Perhaps one barrier is the level of statistical sophistication required to understand the measurements sufficiently well to be convinced of the anticipated benefits of open access. Furthermore, these studies are based on the assumption that quality of research is to be measured by counting citations, whereas most researchers understand perfectly well that citation count is merely an example of one possible way to construct an operational variable for measuring research quality.

Open Access as a Catalyst for Innovation Through Text/Data Mining

Jisc recently published a detailed report outlining and estimating the actual and potential value and benefits of text mining (McDonald & Kelly, 2012). The benefits highlighted in the report include significant increases in efficiency and interdisciplinary scope of literature reviews. The Jisc report lists the lack of the understanding of the potential of text mining as one of the barriers to innovation. The problem with text mining, succinctly summarized in the Nature editorial on this topic, is that "the promise" of text/data mining has not been realized due to restrictive access licenses that do not explicitly permit the necessary copying/annotating that is a part of text mining process (Nature, 2012). Presumably, a complete research literature that was free of restrictive license limitations would make it easier to demonstrate how text/data mining can lead to scientific discovery. Although there are some influential voices advocating for the necessity and benefits of text mining (RCUK Administrator, 2012; Murray-Rust, 2012), they are limited to the biomedical and chemistry fields, where most of the use of text mining has so far been demonstrated. Yet even in the biomedical field itself, the success stories of text mining are not plentiful; a decade of existence of the open access PubMed Central that explicitly facilitates text mining efforts has not resulted in a plethora of articles reporting new discoveries with the use of text mining (Bergman, 2012).

Although we may speculate that the future of scholarly communication requires "machine-actionable" documents (Ayris, 2011), we must not lose sight of the fact that human reading is the more fundamental objective. The Jisc report lists the PDF format of documents as a problem for text mining:

> "The tendency to store papers lodged within institutional repositories as pdfs only further contributes to the problem. XML is the preferred format for text mining" (McDonald & Kelly, 2012).

Several facets of electronic publishing continue to faithfully replicate the conventions, look, feel and usability of print journals, and the PDF version of an article is one such facet, since it is conducive to being printed and read as a typeset paper document. Since the PDF is easy to download or print, it is often well-suited to the needs of readers and scholars. However, while being a handy means to distribute content on the web, it offers limited semantic potential for the creation of new articles. The content within the PDFs is somewhat locked down by the nature of this format.

We should pause here to consider whether the potential of text/data mining to generate new knowledge ought to be a rhetorical strategy used by open access advocates. Should an article's receptivity to machines be a real selling point in the mission to have more openly available scientific literature? The Budapest Open Access Initiative[10] and the Berlin Declaration on Open Access to Scientific Knowledge[11] list many rights of access, including: reading, downloading, copying, distributing, printing, searching, crawling and passing as data to software, making and distributing derivative works. However, it seems obvious that the right to read the content in a human usable format is the most fundamental and universally understood of these rights.

Human beings are the intended audience of journal articles. Since only about 20% of the journal literature is openly available, the vast majority of scientific information still remains unavailable to researchers in many universities, to students worldwide and to the general public. Lack of access to science journals effects researchers and students in both industrialized and developing countries. Even in developed countries, universities invest significant resources to extend their journal collections through consortial licensing, resource sharing and interlibrary loan and document delivery services. The access problem is alive and well in research environments. We do not doubt some of the potential of text mining, but the need for increasing access to human readable research articles ought to remain the main selling point employed by open access advocates.

While some new media researchers, computer scientists and informatics researchers are experiment with text mining technologies, widely used and usable search tools that offer superior search experience to tools such as Elsevier's ScienceDirect[12] are difficult to find. The number of researchers motivated by the promise of innovation through text/data mining is limited compared to the much more significant need for universal access to human-readable research output.

[10] Budapest Open Access Initiative: http://www.soros.org/openaccess/read

[11] Berlin declaration on Open Access to Scientific Knowledge: http://oa.mpg.de/lang/en-uk/berlin-prozess/berliner-erklarung/

[12] ScienceDirect.com: http://www.sciencedirect.com

Mediated Deposit

Mediated Deposit Service

Following an 18-month period of discussion and consultation in all departments and faculties, Concordia University passed the Senate Resolution on Open Access on April 16, 2010. As the text of the resolution indicates, the Senate:

> "...from now on encourages its faculty members to deposit an electronic copy of their refereed research output and creative work in Spectrum, along with nonexclusive permission to preserve and freely disseminate it; and furthermore, in the specific case of any scholarly article accepted for publication in a peer-reviewed journal, from now on requires all faculty members to deposit an electronic copy in Spectrum along with non-exclusive permission to preserve and freely disseminate it. This requirement is not binding in cases where publishers, co-authors or other rights holders disallow such a deposit. Faculty members may also, without prejudice, opt out of the requirement by notifying the University Librarian in writing that their work has appeared, or will appear in another Open Access format; or by citing other factors that currently discourage them from depositing their work in an Open Access repository." (Concordia University, 2010).

After the Concordia University Senate Resolution on Open Access was adopted, we concentrated on improving our services to authors. Indeed, as University Librarian Gerald Beasley noted:

> "We are living with the reality that even a Senate Resolution does not mean faculty members have the time or inclination to deposit their research or creative output in an IR or other OA venue" (Beasley, 2011).

We learned, in the time leading up to the passing of the Resolution, that Concordia's faculty members had three common concerns:

1. The amount of time it took to verify author rights and deposit research in an OA environment;
2. The possible negative impact on scholarly publishing within certain disciplines;
3. The fear that OA implied a loss of authorial control (Beasley, 2011).

While the library could not necessarily address the particular disciplinary concerns of specific faculty members, we were in a position to assist authors with the time and expertise needed to verify author self-archiving rights and to an extent we could address the third concern over perceived loss of authorial control through education and outreach.

 doi:10.2218/ijdc.v8i1.238

A significant component of repository management is the establishment of relationships with faculty members and the provision of services to them. The educational component of repository work is substantial, and addresses a faculty need for more information on the legalities and logistics of depositing. Coupling this information gathering and investigation of publisher policies with mediated deposit and author support has been, for us, the means to begin populating the repository. Thus the Senate Resolution was as much of a mandate for faculty members as it was for those implementing and developing Spectrum as a campus service.

Beginning in Spring 2011, we began to gauge the extent of faculty publishing in journal articles. If we were to assist in the university's development of Spectrum as a service, we needed to have a sense of the research output. We developed saved searches and email alerts in various publisher and aggregator databases in order to begin to understand faculty publication patterns. This would allow us to observe which publisher policies would be the most important to know. It was never assumed that the system of monitoring search alerts would yield a comprehensive account of total publishing output at the University, since not all publisher websites have easily retrievable and exportable search results sorted by author affiliation, and aggregator databases may not index all publications in which a university's authors publish.

We regularly receive publisher and database search alerts through email, then input the citations into RefWorks, a citation management software. Duplicate citations are removed. The citations are organized in folders to denote the author's department and the appropriate version of the article that is eligible for deposit. This preparatory work, performed by the library, allows us to develop a proactive outreach program to encourage and facilitate faculty deposits in Spectrum. An individual citation, including the multiple folders in which it is grouped, allows Spectrum editors or subject librarians to contact faculty members at or near the time of publication with specific information regarding their options for deposit. Most faculty members respond to this strategy, and appreciate the service.

The method for monitoring faculty publications and recruiting Spectrum content is summarized with the following steps:

1. Create and receive regular email alerts from Saved Searches & Email Alerts features in publisher and aggregator databases;
2. Import alerts into RefWorks as they arrive (usually on a daily or weekly basis) using "Import/Export" functionality at the aggregator/database;
3. Remove duplicate citations in RefWorks;
4. Organize and group citations using "folders" in RefWorks. Group by department and publisher self-archiving policy, noting which version can be deposited and whether there is an embargo period;
5. Using email templates for different deposit scenarios in the previous step, send emails to authors asking for permission to proceed and request a file for upload;
6. If the author supplies the file, the editor uses the DOI import to make the deposit.

Impact of Mediated Deposit Service

From Spectrum's inception, Spectrum editors were willing to make deposits on behalf of faculty members, as we knew from the literature that authors are often too busy to make deposits themselves or are reluctant to investigate their publisher's policies on self-archiving. However, in the earliest months of depositing it was mostly authors who were doing their own deposits. When the resolution was passed in spring 2010, deposits continued to increase but the library also began directing more human resources towards assisting faculty members with deposits. One year later, in spring 2011, we began monitoring search alerts and reaching out to faculty members whose publications were eligible for deposit.

As we can see in Figure 1, during the first year of our more directed outreach to authors, the number of mediated deposits began to surpass author self-archiving. This is undoubtedly a result of our increased efforts to assist faculty members, though as the graph relates, self-depositing is still occurring. Some authors complete their own deposits after receiving an outreach email, while others simply forward the post-print to a Spectrum editor.

In our view, this hybrid method for populating Spectrum allows authors to participate in the Senate Resolution in a way that suits their goals, time availability, comfort with the technology and willingness to explore their publisher's policies on self-archiving. For some authors, the information provided in the outreach email about their self-archiving option is the key impetus to deposit, while for others, it is the fuller service available in which a librarian makes the deposit. By offering the service, we attempt to address and mitigate both information and time shortages on the part of busy researchers.

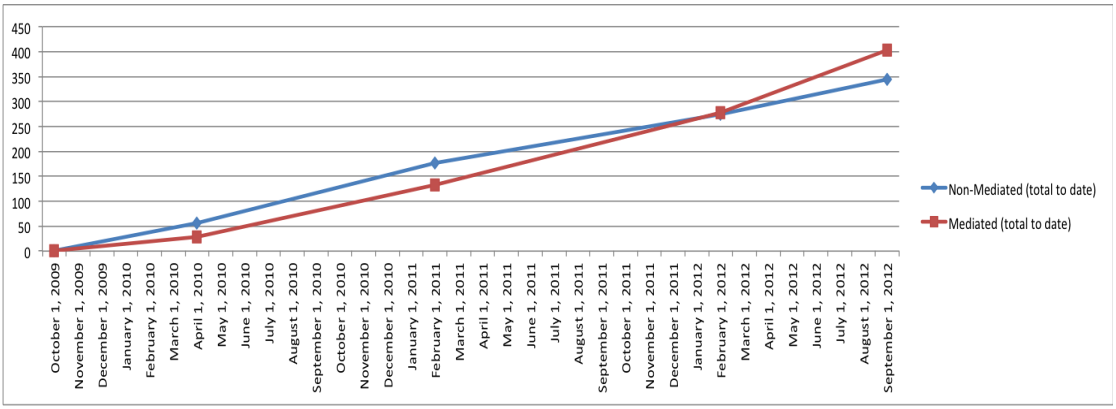

Figure 1. Total deposits of non-thesis materials (articles, conference presentations, etc.) to Spectrum: Concordia University Research Repository.

Conclusions

Open access offers wider distribution of research articles, thereby increasing access to research, which can increase its impact. It can increase the visibility, readership, use and uptake of research, but can it really improve the quality of the science in the

 doi:10.2218/ijdc.v8i1.238

articles? If it could, what would be the best metric for gauging the research quality of openly available articles vs. articles not openly available?

The marketing of open access as a stepping stone towards Artificial Intelligence on the Web is inappropriate because it relies on speculative assumptions about the benefits of technologies that have yet to be tested for human-centered usability and value. Automation has to serve human decision-making – human peer review is the central pillar of quality control in scholarly publishing. In order for peer review to work well, it makes sense for reviewers to have ready access to the works cited in the article under review, as well as the broadest possible range of scientific literature. Thus, the argument for open access is sufficiently justified by the fact that the quality of peer review is dependent on human reviewers having universal access to research results. The best possible environment for a scholarly information ecosystem is one that is open.

It is exciting to consider the possibility of a corpus of scholarly research that is openly available, fully-searchable, well indexed on the web and capable of being itself a ground of inquiry for analysis and compilation of research results. However, we believe that promoting open access as a means to creating a computer-readable storehouse of data that will lead to new knowledge and discoveries generated by computer algorithms is too speculative as an incentive for researchers to adopt author self-archiving or publish in toll-free journals.

Mandates from funders and universities worldwide will likely be the only mechanism that will encourage authors to place an open access copy of their work in a repository. Good intentions, altruism and a sympathetic stance with regards to open access and its ideals have not yet generated enough voluntary deposits compared to the total number of research articles published per year. It would appear that the case for open access has not yet been made in a way that truly resonates with academics. Although many advocates for better public access to scholarly and scientific research view the deposit of a post-print of a published article as a simple transaction taking only minutes, it has not been internalized and adopted as a regular practice for academics.

Beyond mandates, we feel that the way to encourage authors to embrace open access is to provide them with systems and services that make it easy to share their work on the web. Easy-to-use repository interfaces and helpful assistance from repository managers and staff, combined with the ability to shorten the deposit process through proven technical means, such as a DOI import or using the SWORD protocol, are the means by which we can improve the likelihood of mandate compliance. Importantly, academics require assistance with the necessary interpretation of publisher policies or the sometimes necessary task of obtaining permission from a publisher, as both activities are required before a deposit can be made. The universal acceptance and uptake of open access, particularly author self-archiving, is in fact a significant and complex socio-technical behavior change, with various stakeholders from the academy and publishing alike needing to adjust practices in concert with one another.

John Willinsky suggests that “there will come a point where open access to research and research is commonplace” and he speculates that “the advantages of greater accessibility will translate into an improved reputation for research as a whole”

(Willinsky, 2010). These remarks take into account the increased usefulness and adoption of research that the digital age, and open access in particular, can stimulate. Willinsky also suggests, "for most faculty members, the digital era has changed little except the ease with which they access the literature, and the speed with which they correspond with colleagues and journals."

Indeed we have not yet witnessed the massive transformation envisioned (i.e. universally mandated green author self-archiving in open access repositories) by one of Open Access's leading advocates, Steven Harnad. Although "spontaneous self-archiving" of post-prints in institutional repositories may be dearly wished for by those engaged in open access, such a dramatic change in researcher behavior suggests that no amount of clearing up misunderstandings about open access will persuade scholars to engage in a new "spontaneous" activity. There is also an inherent tension in proposing that universal mandates will lead to "spontaneous self-archiving." A spontaneous action is one that is "arising or proceeding entirely from natural impulse, without any external stimulus or constraint; voluntary and of one's own accord" (OED, 2012). An open access mandate is, by definition, an external stimulus. Can a voluntary action be mandated?

Nonetheless, until funder and institutional mandates become more commonplace or behavior shifts significantly amongst researchers, those involved with open access repositories should offer advocacy and services that suit the disciplinary culture of the researchers at their institutions, clearly and effectively demonstrate the usage and citation impact of OA by offering an institutional repository that measures use, and do as much as possible to assist researchers in making their work available on the Web. Stevan Harnad writes that the cure for researcher and author reluctance to self-archive is the open access mandate. He writes:

> "...in the case of providing OA to peer-reviewed research, it has turned out that the way to 'change behaviour' is to mandate it" (Harnad, 2011).

In our experience, mandating a behavior change is but one step. Our experience operating a repository at a university with a mandate in place has shown that there are various reasons why a mandate might not beget a more complete participation. A significant resource commitment, shared by the Library and any number of relevant campus units, is required in order to create the conditions that bring about regular self-archiving behavior. We have learned that the faculty concerns over publisher policies and lack of time are not necessarily resolved by a mandate. However, a mandate motivates a University to develop and refine the outreach strategies and services that will help their particular university milieu, which will necessarily exist in various overlapping socio-technical and disciplinary contexts.

References

Ayris, P. (2011). Conference report: OAI7 – Raising the bar for current developments in scholarly communication, a personal view. *LIBER Quarterly: The Journal of the Association of European Research Libraries 21*(1). Retrieved from http://liber.library.uu.nl/index.php/lq/article/view/8013/8356

Beasley, G. (2011). The senate resolution on open access at Concordia University. *Feliciter 57*(3), 101-102.

Bergman, C. (2012). *Why are there so few efforts to text mine the open access subset of PubMed Central?* Retrieved from http://caseybergman.wordpress.com/2012/03/02/why-are-there-so-few-efforts-to-text-mine-the-open-access-subset-of-pubmed-central/

Bjork, B.C., Roosr, A. & Lauri, M. (2008). Global annual volume of peer reviewed scholarly articles and the share available via different open access options. *Proceedings of the International Conference on Electronic Publishing, 12,* 178-186. Retrieved from http://elpub.scix.net/cgi-bin/works/Show?178_elpub2008

Brinn, T., Jones, M. J. & Pendlebury, M. (2000). Measuring research quality: Peer review 1, citation indices 0. *Omega, 28*(2), 237-239. doi:10.1016/S0305-0483(99)00048-1

Brody, T., Harnad, S. & Carr, L. (2006). Earlier web usage statistics as predictors of later citation impact. *Journal of the American Society for Information Science & Technology, 57*(8), 1060-1072. doi:10.1002/asi.20373

Budapest Open Access Initiative. (2012). *Ten years on from the Budapest Open Access Initiative: Setting the default to open.* Retrieved from http://www.soros.org/openaccess/boai-10-recommendations

Bunge, M. (2004). Clarifying some misunderstandings about social systems and their mechanisms. *Philosophy of the Social Sciences, 34*(3), 371-381. doi:10.1177/0048393104266860

Bunge, M. (2003). *Emergence and convergence: Qualitative novelty and the unity of knowledge*. Toronto: University of Toronto Press.

Bunge, M. (1998). *Social science under debate: A philosophical perspective*. Toronto: University of Toronto Press.

Concordia University. (2010). *Concordia University senate resolution on open access.* Retrieved from http://library.concordia.ca/research/openaccess/SenateResolutiononOpenAccess.pdf

Davis, P.M. (2011). Open access, readership, citations: A randomized controlled trial of scientific journal publishing. *The FASEB Journal*. doi:10.1096/fj.11-183988

Davis, P.M. (2010). Does open access lead to increased readership and citations? A randomized controlled trial of articles published in APS journals. *The Physiologist, 53*(6), 197, 200-1.

Davis, P.M. (2009). Open access: Increased citations not guaranteed. *Science, 325*(5938), 266. doi:10.1126/science.325_266a

Foster, N.F. & Gibbons, S. (2005). Understanding faculty to improve content recruitment for institutional repositories. *D-Lib Magazine, 11*(1), 1-1. doi:10.1045/january2005-foster

Fry, J., Oppenheim, C., Probets, S., Creaser, C., Greenwood, H., Spezi, V. & White, S. (2009). PEER behavioural research: Authors and users vis-à-vis journals and repositories. LISU: Loughborough University. Retrieved from http://www.peerproject.eu/fileadmin/media/reports/Final_revision_-_behavioural_baseline_report_-_20_01_10.pdf

Gargouri, Y., Hajjem, C., Larivière, V., Gingras, Y., Carr, L., Brody, T. & Harnad, S. (2010). Self-selected or mandated, open access increases citation impact for higher quality research. *PLoS ONE, 5*(10), 1-12. doi:10.1371/journal.pone.0013636

Gentil-Beccot, A., Mele, S., & Brooks, T. (2010). Citing and reading behaviours in high-energy physics. *Scientometrics, 84*(2), 345-355. doi:10.1007/s11192-009-0111-1

Nature. (2012). Gold in the text? *Nature 483*(7388). doi: 10.1038/483124a

Guedon, J. (2003). Open access archives: From scientific plutocracy to the republic of science. *IFLA Journal, 29*(2), 129-140. doi:10.1177/034003520302900204

Harnad, S. (2000). The invisible hand of peer review. *Exploit Interactive* 5. Retrieved from http://www.exploit-lib.org/issue5/peer-review/

Harnad, S., Brody, T., Vallières, F., Carr, L., Hitchcock, S., Gingras, Y. & Hilf, E.R. (2004). The access/impact problem and the green and gold roads to open access. *Serials Review, 30*(4), 310-314. doi:10.1016/j.serrev.2004.09.013

Harnad, S. (2008). *Institutional repositories vs subject/central repositories - open access archivangelism.* Retrieved from http://openaccess.eprints.org/index.php?/archives/414-guid.html

Harnad, S. & McGovern, N. (2009). Topic 4: Institutional repository success is dependent upon mandates. *Bulletin of the American Society for Information Technology and Science 35*, 27–31. doi:10.1002/bult.2009.1720350410

Harnad, S. (2010.) Gold open access publishing must not be allowed to retard the progress of green open access self-archiving. *Logos: The Journal of the World Book Community (21)*3-4, 86-93. doi:10.1163/095796511X559972

Harnad, S. (2011). Open access to research: Changing researcher behavior through university and funder mandates. *JEDEM Journal of Democracy and Open Government 3*(1), 33-41.

IEEE. (2011.) Peer review notice. *CS Digital library.* Retrieved from http://www.computer.org/portal/web/csdl/peerreview

Kennan, M.A. (2011). Learning to share: Mandates and open access. *Library Management, 32*(4), 302-318. doi:10.1108/01435121111132301

Kim, J. (2010). Faculty self-archiving: Motivations and barriers. *Journal of the American Society for Information Science & Technology, 61*(9), 1909-1922. doi:10.1002/asi.21336

Lawrence, S. (2001). Free online availability substantially increases a paper's impact. *Nature, 411*(6837), 521-521. doi:10.1038/35079151

Linde, P. (2010). Beyond OA-policy. *ScieComInfo: Nordic-Baltic Forum for Scientific Communication 6*(2). Retrieved from http://www.sciecom.org/ojs/index.php/sciecominfo/article/view/3623

Lynch, C.A. (2003). Institutional repositories: Essential infrastructure for scholarship in the digital age. *ARL, 226.* Retrieved from http://www.arl.org/storage/documents/publications/arl-br-226.pdf

McDonald, D. & Kelly, U. (2012). The value and benefits of text mining. London: Jisc. Retrieved from http://www.jisc.ac.uk/publications/reports/2012/value-and-benefits-of-text-mining.aspx

Morris, S. & Thorn, S. (2009). Learned society members and open access. *Learned Publishing, 22*(3), 221-239. doi:10.1087/2009308

Murray-Rust, P. (2012). Towards a manifesto on open mining of scholarship. Retrieved from http://blogs.ch.cam.ac.uk/pmr/2012/05/01/towards-a-manifesto-on-open-mining-of-scholarship/

Open Citation Project. (2012). *The effect of open access and downloads ('hits') on citation impact: A bibliography of studies.* Retrieved from http://opcit.eprints.org/oacitation-biblio.html

Open Access Directory. (2011). *Declarations in support of OA.* Retrieved from http://oad.simmons.edu/oadwiki/Declarations_in_support_of_OA

Open Access Directory. (2011). *Statements by learned societies and professional associations.* Retrieved from http://oad.simmons.edu/oadwiki/Statements_by_learned_societies_and_professional_associations

RCUK Administrator. (2012). The benefits of Open Access. Retrieved from http://blogs.rcuk.ac.uk/2012/08/10/the-benefits-of-open-access/

Sale, A. (2006). The acquisition of open access research articles. *First Monday, 11*(10).

Sale, A. (2006). Comparison of content policies for institutional repositories in Australia. *First Monday, 11*(4).

Schonfeld, R.C. & Housewright, R. (2010). Faculty survey 2009: Key strategic insights for libraries, publishers, and society. Ithaka S+R. Retrieved from http://cyber.law.harvard.edu/communia2010/sites/communia2010/images/Faculty_Study_2009.pdf

Seglen, P.O. (1997). Citations and journal impact factors: Questionable indicators of research quality. *Allergy, 52*(11), 1050-1056. doi:10.1111/j.1398-9995.1997.tb00175.x

Sieber, J.E. (2006). Quality and value: How can we research peer review? *Nature,* doi:10.1038/nature05006

OED Online. (2012). *Spontaneous*. Oxford University Press. Retrieved from http://www.oed.com/view/Entry/187385

Suber, P. (2010.) Thoughts on prestige, quality, and open access. *Logos: The Journal of the World Book Community 21*(1-2), 115-128. doi: 10.1163/095796510X546959

Suber, P. (2011). Knowledge as a public good. Retrieved from http://www.arl.org/sparc/publications/articles/knowledge-public-good.shtml

Swan, A. & Brown, S. (2005). *Open access self-archiving: An author study.* Truro, England: Key Perspectives. Retrieved from http://www.keyperspectives.co.uk/openaccessarchive/reports/Open%20Access%20II%20(author%20survey%20on%20self%20archiving)%202005.pdf

Swan, A. (2010). *The open access citation advantage: Studies and results to date.* Retrieved from http://eprints.ecs.soton.ac.uk/18516/

van Rooyen, S., Godlee, F., Evans, S., Smith, R. & Black, N. (1998). Effect of blinding and unmasking on the quality of peer review. *JAMA: Journal of the American Medical Association, 280*(3), 234-237. doi:10.1001/jama.280.3.234

Wagner, A.B. (2010). Open access citation advantage: An annotated bibliography. *Issues in Science & Technology Librarianship,* 60, 2-2. doi:10.5062/F4Q81B0W

Ware, M. (2008). Peer review: Benefits, perceptions and alternatives. Summary Paper 4. London: Publishing Research Consortium. Retrieved from http://www.publishingresearch.net/PeerReview.htm

Willinsky, J. (2010.) Open access and academic reputation. *Annals of Library and Information Studies 57* (3), 296-302. Retrieved from http://nopr.niscair.res.in/handle/123456789/10242

Xia, J. (2011). An anthropological emic-etic perspective on open access practices. *Journal of Documentation, 67*(1), 75-94. doi:10.1108/00220411111105461