

Faculty self-archiving attitudes and behavior at research universities

A Literature Review

Molly Kleinman

ED622: Proseminar in Higher Education

December 9, 2011

Note: This literature review was written as the final paper in a doctoral seminar at the Center for the Study of Higher Education at the University of Michigan taught by Dr. Deborah Carter. As in any class assignment, there were constraints such as page limits and required sections on things like methodology and theoretical frameworks. In particular, students were required to use only peer-reviewed literature their reviews. As a result, a handful of widely cited and rigorous studies on self-archiving were excluded from this study because they appear only in gray literature or in venues without formal peer review, and therefore this literature review is not comprehensive. In short, yes yes, I know I've left out the work of Alma Swan and several reports from CSHE, among others. It was with regret.

After winning the Nobel Prize for co-discovering the structure of DNA, Francis Crick commented that despite the risk that competitors could use it to their advantage, the X-ray which led to his discovery had to be shared publicly because “communication is the essence of science” (quoted in Garvey, 1979). Scholars must publish and distribute their research in order to participate in an ongoing conversation with their peers, and with the researchers that came before and those that will come after. New research depends upon the findings of prior research; scholarship requires publication. The basic methods by which researchers communicate their findings to colleagues – peer reviewed journal articles and monographs – have remained remarkably consistent over the last century, even though the economics of the system changed dramatically with the rise of the internet (Bergstrom, 2001). Control over scholarly work has migrated from scholars and non-profit journals into the hands of a few large corporations, resulting in reduced access to scholarship even for researchers at the wealthiest institutions, and causing what has become known as “the scholarly communications crisis” (Bergstrom and Bergstrom, 2006).

The crisis is exemplified by a study of ecology journals by Carl and Theodore Bergstrom which compared the pricing of journals published by for-profit publishers with those published by scholarly societies and other non-profit publishers (2006). The Bergstroms found that comparing both price per page and price per recent citation, for-profit journal subscriptions cost five times as much as non-profit journals in the same field. This is true even when controlling for the quality of the journal, using citation rates as a proxy. They go on to demonstrate that high prices lead to decreased access by showing that a journal’s circulation is closely connected to its price. The higher the cost of a journal, the fewer libraries subscribe to it, and therefore fewer scholars are able to access it through their institutional affiliations. Despite these inefficiencies,

the Bergstroms observe that faculty continue to publish in expensive, for-profit journals because that is where their colleagues publish, and also what their colleagues read. In short, even though for-profit journals are incurring significant strain on the scholarly communication system as a whole by limiting access, there are few incentives for faculty to change their behavior.

One attempt to ameliorate the scholarly communications crisis has come in the form of the open access (OA) movement. The open access movement advocates for free, online access to peer reviewed scholarship (Suber, 2010). It is grounded in two basic arguments: 1) The internet makes possible free and instantaneous distribution of knowledge, and so academia should avail itself of the efficiencies afforded by the network; and 2) the scholarly communication system is broken; faculty write scholarly articles for free, based on research that was often funded by taxpayers, and those articles should be freely available to the public (Willinsky, 2002; Wellen, 2004). Though both arguments are compelling, much of the discussion has centered on the latter, and the potential for open access to repair or replace the current system (Bailey, 2006).

The open access movement has coalesced around two ways for a work to be open access. Either the author publishes an article in a traditional journal and posts a copy on a publicly available online repository or website (self-archiving), or the author chooses to publish in one of the growing number of open access journals, which are made freely available online from the moment of publication (open access publishing) (Bailey, 2006). Self-archiving has the advantage of permitting faculty to continue publishing in whichever journals they choose, while enabling broader access to their work. It is an option for all scholars, regardless of discipline, as long as they have access to the internet. Open access journals serve as direct competition for traditional for-profit publishers, but a reputable open access journal must exist in the author's field in order for open access publishing to be a viable option (Park and Qin, 2007).

Both self-archiving and open access publishing are relatively young concepts, only as old as the internet. They have experienced rapid growth in the last 15 years, as the web has become more widely available and network speeds have increased (Suber, 2010). As a result, the study of open access practices among research faculty is quite new, and methodologies and conceptual frameworks for understanding them are still under development. Researchers are just starting to uncover the incentives and influences affecting faculty self-archiving and open access publishing behavior, and little work has been done to synthesize these results into a comprehensive understanding of what impact the open access movement is having on the scholarly communications system as a whole.

This literature review investigates the self-archiving attitudes and behavior of faculty at research institutions, primarily in the United States. Because some form of self-archiving is an option for all scholars regardless of discipline, and it may therefore have the potential to make a more immediate impact on the scholarly communications system, this review focuses on self-archiving behavior and does not investigate open access publishing. The significance of this review represents the need to uncover the motivations and drivers behind faculty willingness to share their work online, and to identify what is still unknown about faculty attitudes towards open access. The central questions this review seeks to address include:

1. What do we know about the attitudes and behavior of research faculty with regard to self-archiving their scholarly work?
2. What theoretical and conceptual frameworks are being used to understand those attitudes and behavior?
3. Has the existing research uncovered the influences and incentives that affect the decision to archive work in open repositories? If so, what are they?

4. What are the gaps in our understanding of this topic, and where should we focus future research?

The following section investigates the varied definitions researchers use for key terms like “open access” and “self-archiving,” in order to come to a stable understanding of the terminology used throughout the review. The subsequent two sections review studies on faculty attitudes towards self-archiving and faculty self-archiving behavior respectively, followed by a brief discussion of the impact of public access mandates. The review concludes with overviews of the prominent methodologies, conceptual frameworks, and limitations prominent in the field, and implications for future research.

Defining “Open Access,” “Self-Archiving,” and “Repository”

The competing definitions of open access are essentially multiple factions in the same camp. All sides agree that to be considered open access, literature must be freely available online in full text for access by anyone with an internet connection (Bailey, 2006). No one disputes that open access can be achieved in two kinds of venues, open access journals and open access repositories. Though monographs and data are increasingly included under the open access umbrella, all of the studies in this review focus on peer-reviewed journal articles, which are the type of work mostly commonly discussed by open access advocates (Joseph, 2008). Contention arises around one issue: copyright barriers. Definitions rooted in the Budapest Open Access Initiative from 2001 state that in addition to free availability, OA literature must be released under open copyright licenses that allow unrestricted re-use without requiring permission or payment (Suber, 2010). Proponents of these definitions argue that simply making something available for free on the internet is not sufficient for “open access”; works must also be free to adapt, reuse, and build upon. Others, particularly librarians and other practitioners, believe that

free access online is sufficient to apply the “open” label (Bailey, 2006).

These two flavors of access are often termed “gratis OA” and “libre OA” (Suber, 2010; Salo, 2010) to draw the distinction between works that are only free to access and works that are free both to access and to reuse. Every article in this review uses “open access” to mean gratis OA. Most of the authors do not provide their own definitions of OA, but when they do, it is brief and in passing: “mak[ing] work freely available on the web” (Watson, 2007, p. 226) and “depositing articles published in non-OA journals into OA archives” (Kim, 2010, p. 1909). Though some of the articles investigate faculty attitudes towards copyright control (Gadd, Oppenheim, and Proberts, 2003; Kim 2010; Lercher, 2008), none require freedom from licensing restrictions to classify a work as open access.

This happy consensus among the authors reviewed here does not extend to definitions of “self-archiving,” however. In his chapter synthesizing the debates about key terms in open access, Bailey defines self-archiving this way: “When authors make their articles freely available in digital form on the Internet, they are said to be ‘self-archiving’ them. These articles can be either ‘preprints’ or ‘postprints’ ” (2006, p. 19). Many of the articles in this review provide more nuance on one or more elements of this basic definition, with contrasting approaches. Some allow that free availability anywhere on the internet qualifies as self-archiving (Covey, 2009; Gadd, Oppenheim, and Proberts, 2003; Kim 2007; Kim 2010), while others require deposit in an actual archive, either an institutional or disciplinary repository (Lawal, 2002; Lercher 2008; Sale 2006; Watson, 2007; Xia 2008; Xia and Sun, 2007; Kim, 2011). Many of the articles treat a work as self-archived regardless of who posted it online, while a handful distinguish between deposits performed by the author herself (self-archiving) and those performed on her behalf (mediated deposit) (Watson, 2007; Xia 2007; Xia and Sun 2007).

Wrapped up in this confusion are the terms pre-print, post-print, and e-print, and the classes of material they aim to represent. In general, pre-print refers to articles that are intended for publication but are not published yet, post-print refers to articles that have already been published, and e-print is used to mean different things by different people. Most of these studies do not distinguish between pre- and post-publication articles. However, Lercher (2008) focuses exclusively on unpublished material, while Lawal (2002) claims to study only e-prints, which he defines as pre-publication articles in electronic form, but he investigates repositories that generally hold both pre-and post-publication materials, calling them all “e-prints”. While these details may be relatively minor, taken together they result in different methods of measuring self-archiving that may make it difficult to draw broad conclusions from this body of literature.

In contrast, the definitions of repository types throughout the articles in this review, both institutional and subject based, are remarkably consistent with each other. The definitions of “institutional repository (IR)” all match the one Shreeves and Cragin (2008) use in their introduction to a special issue of *Library Trends* about IRs. They define an institutional repository as a set of services and technologies that provide the means to collect, manage, provide access to, disseminate, and preserve digital materials produced at an institution. Note that this definition does not specify the kind of digital materials; though journal articles are the dominant content type in many IRs, most now hold a wide range of objects, including data, gray literature, and educational resources (Shreeves and Cragin, 2008). Throughout the literature reviewed here, the terms “subject repository (SR)” and “disciplinary repository (DR)” are used interchangeably to describe repositories based on subject matter and that are not institution specific. For consistency, I will use disciplinary repository throughout this review.

Criteria for Inclusion

Articles were selected for inclusion in the review if they met three criteria: 1) Their primary focus was on research faculty behavior or attitudes with regard to self-archiving practices; 2) they were empirically grounded; and 3) the article was published after 2001. Studies concerned principally with the systems or technologies underlying institutional or subject repositories were not included. Because very few universities had repositories prior to 2001, this review is limited to studies conducted after IRs became more widely available (Shreeves and Cragin, 2008). The studies generally fell into two categories: investigations of faculty attitudes towards self archiving which tended to rely on surveys, interviews, or a combination of the two, and investigations of faculty self-archiving behavior, which generally were more quantitative in nature and drew upon server and collections data directly from repositories.

Faculty Attitudes Regarding Self-Archiving

Comparing Archivers with Non-Archivers

Several studies investigated the difference between faculty who self-archive and those who do not. Gadd, Oppenheim, and Proberts conducted an international author survey as a part of the UK-based RoMEO Project, which sought to develop a comprehensive understanding of the legal and technical requirements for establishing a successful open access infrastructure (2003). The survey focused heavily on the impact of copyright and plagiarism concerns on faculty attitudes toward archiving. It found that confusion about what rights belong to the author and what rights belong to the publisher prevented faculty from archiving, results that mirror those of Kim (2010).

Participants in the RoMEO study were recruited using websites and listservs, and there were a total of 542 respondents, a third of which were from the UK. The sample was skewed

towards the hard sciences, and 58% of the participants had self-archived, a high number that strongly suggests a selection bias (Gadd, Oppenheim, and Proberts, 2003). Survey questions asked what kinds of permissions, such as photocopying, annotating, and excerpting, faculty would be comfortable granting to users of work that they posted online. They found that overall, the protections granted to scholarly work by copyright law exceed the level of protection desired by scholars. The survey also asked where authors had self-archived their work; as in other studies (Kim, 2010), the most common response was on a personal or departmental website.

Though much of the article reports straight percentages of respondents, Gadd et al. used chi squared tests to determine whether there was a statistically significant difference between the general concerns of archivers and non-archivers. There was. The most common response among non-archivers was not actually a copyright concern, but a fear that if they archived an article online prior to publication, they would be unable to find a publisher subsequently. The most common response for archivers was that they had no concern. However, the second most common concern was the same for both groups: fear of breaking agreements with publishers. This finding supports the need to understand the impact of publisher policies on self-archiving, a question investigated in more depth by Covey (2008) and Antelman (2006), discussed later in this review. Given the widespread confusion and fear surrounding copyright and self-archiving in the academy (Willinsky, 2002), this study offers some valuable detail about the sources of those concerns. However, its broader relevance is limited by the self-selected nature of the sample, the lack of conceptual framework, and the narrow application of statistical methods.

Jihyun Kim conducted a series of three studies that focused on the differences in attitudes between faculty who self-archive and those who do not to develop a set of consistent factors that influence self-archiving behavior (Kim 2007; Kim 2010; Kim 2011). These three articles include

detailed methods sections, thoughtful application of theoretical frameworks, and rigorous use of statistical analysis, all of which set them apart from many of the studies that follow in this review. The first study involved a pilot survey of 31 faculty at a single research institution, while the latter two used the same large set of survey and interview data collected from 17 Carnegie doctorate-granting universities in the United States. The survey and interview instruments for the multi-institutional studies were developed based on a literature review and on findings from the first study, and the final study refines the factor analysis initiated in the second study. Using this iterative approach, Kim has developed a seemingly robust and compelling set of seven factors that significantly influence faculty attitudes towards self-archiving: copyright concerns, age, and additional time and effort were negatively related, while academic reward, altruism, self-archiving culture, and technical skills were positively related.

For the latter two studies, Kim sampled faculty from 17 doctorate-granting universities that all used the same software platform (DSpace) for their institutional repositories, in order to reduce the variability in experience with IRs across institutions (2007). However, it should be noted that Kim used a very broad definition of self-archiving which included posting on a personal or departmental website; most of the self-archivers in her study did not actually use their IRs. Kim surveyed a random sample of 684 faculty, drawn from two separate pools of those who had self-archived and those who had not, and conducted follow-up interviews with 41 of the survey participants; the sample was cross-disciplinary, and included faculty at all levels, from assistant professor through full professor.

Kim grounded her studies in two related conceptual frameworks: the socio-technical interaction network (STIN) model and social exchange theory. The STIN framework was designed to explain the interactions between the social and technical elements of networked

communications forums (Kling, 2003); in the case of IRs, Kim applied it to develop an understanding of the relationship between the faculty who deposit their work into a repository, and the database and other technical infrastructure that makes the repository work (2007). Social exchange theory describes actions that elicit rewards from others, and is intended to encompass a broad array of exchanges beyond the monetary (Emerson, 1976). Kim used social exchange theory to examine the sharing of information through an economic lens, and to address the issue of incentives, perceived costs and benefits, and contextual factors that influence information sharing behavior (2010).

In the 2010 study, Kim employed ordinary least squares regression and factor analysis to tease out the variables with the strongest impact on self-archiving attitudes, and found that altruism had the greatest effect, followed by perceived archiving culture in the discipline, and copyright concerns. The 2011 study used logistic regression to identify trust as another key factor. Faculty who did not trust others to use their work properly and refrain from plagiarism were less likely to self-archive in any form. Rank was found to have no significant effect, though age did. It is interesting to note that a subject's discipline alone was not a predictor of positive attitudes toward self archiving; it depended upon the subject's perception of the archiving culture in her discipline. Jingfeng Xia suggested that discipline has no effect on self-archiving (2008); Kim's finding that influence comes from a subject's perception of disciplinary culture offers a middle ground between no disciplinary effect and a strong disciplinary effect.

Understanding Non-Participation in Institutional Repositories

Two case studies broached the question of negative faculty attitudes towards self-archiving by seeking the reasons for non-participation in IRs (Watson, 2007; Davis and Connolly, 2007). Watson conducted 21 interviews with faculty at Cranfield University in the UK with the

aim of determining what might motivate them to deposit in the university's IR, called QUEPrints (2007). Participants were selected from a range of disciplines, and some were chosen either because they had already made multiple deposits in QUEPrints or because they held influential positions such as department head. Watson found that 57% of faculty had heard of QUEPrints, and only 43% knew what it was, suggesting that a primary reason for non-participation was simple lack of awareness. When faculty were asked about the theoretical reasons they might participate in QUEPrints, even if they had not done so already, improving access was the most common response. Of those who had deposited in QUEPrints, the most common reasons were that someone, usually a librarian, had asked them to deposit, and that they wanted to increase the visibility of their work. This study, conducted by a librarian at Cranfield University, was very inward facing, and focused mostly on implications for future practice at Cranfield without making suggestions for broader application of the results or suggestions for additional research.

In a similar case study, Davis and Connolly (2007)¹ used a mixed methods approach to understand non-use of Cornell's DSpace-based repository. First they collected deposit data from the Cornell institutional repository, then they collected similar deposit data from seven other research institutions with DSpace installations for comparison, and finally they conducted eleven interviews with Cornell faculty from a range of disciplines. In contrast with Kim's finding that altruism has a significant impact on self-archiving (2010), Davis and Connolly found that when they inquired about improving access as a motivating factor, the majority of faculty replied that

¹ It must be noted that this article was published in D-Lib Magazine, a prominent journal on digital library issues that is not peer reviewed. Of the seven articles included in this review that were published after 2007, five cite Davis and Connolly (Kim, 2010; Kim 2011; Lercher, 2008; Xia, 2008; Duranceau, 2008). All of these corroborate some portion of Davis and Connolly's findings, while Kim (2010) and Xia (2008) both construct their research models in part based on the Cornell results. Therefore, I would argue that its inclusion here is valuable enough to set aside the peer review requirement in this case.

they did not think about access with regard to their own work. Only a Romance Studies professor, who works extensively with colleagues in South America, gave any consideration at all to access issues. Meanwhile, factors that did affect Cornell faculty included technical confusion about how to use the IR, copyright concerns, and fear of plagiarism; similar concerns have been identified by other studies in this review (Kim 2010; Gadd et al. 2003). As at Cranfield University (Watson, 2007) and the University of Rochester (Bell et al., 2005), one crucial reason that faculty did not use Cornell's IR is that they had not heard of it.

The purpose of the Cornell study was to uncover the reasons that faculty were not using Cornell's institutional repository, and hence other forms of open access participation are discussed only as a possible explanation for non-participation in the IR. Of the 11 faculty interviewed, all but the two humanities scholars shared their work online through personal or departmental websites, and some also used disciplinary repositories. This is an unusually high rate of self-archiving activity compared to the other studies in this review (Kim 2010; Watson, 2007; Covey, 2009), and it is not clear whether or not this was intentional. The article does not state how the subjects were selected, except to say that the researchers sought to include faculty from a wide range of disciplines, from different stages in their careers, and with balanced representation of the sexes. It does not appear that the sample was randomly selected, and there is no attempt to establish statistical significance. Even so, it would have been valuable to see this result plumbed in more depth. If the ultimate goal of institutional repositories is to improve access to scholarship (Shreeves and Cragin, 2008), a finding that most scholars already provide open access to their work in other ways could be a blow for IR proponents.

Unpublished Work and Self-Archiving

Three studies focused on the use of IRs for distributing unpublished or gray literature

(Lercher, 2008; Lawal, 2002; Bell, Foster, and Gibbons, 2005). Lercher conducted a survey of 72 faculty at Louisiana State University (LSU), which did not have an IR at the time of the study, with two primary goals: to identify general faculty attitudes and desires regarding self-archiving services, and to determine whether faculty who believed that they or their colleagues had valuable unpublished work would be significantly more inclined to favor both institutional and disciplinary repositories (2008). Using logistic regression, Lercher concluded that faculty who had valuable unpublished work or believed that others did were indeed more likely to answer yes to questions such as, “Would you post this work to your own website?” and “Would you deposit this work in an IR?” Lercher also hypothesized that faculty would demonstrate a preference for disciplinary repositories over institutional repositories, but the data did not support that assertion. The article raised several questions about variations by discipline and the influence of searching behavior on self-archiving, but it was not clear that the data supported any conclusions. The analysis, discussion, and conclusion sections of the article are all very brief, and the results of the regression are offered in large tables with little additional explanation, which make it difficult to determine which results are meaningful and limiting the broader applicability of the study.

Bell et al. approached a similar question using a participant observation technique with a small sample of faculty subjects across several disciplines at the University of Rochester, accompanied by interviews with subject librarians (2005). Unfortunately, the findings presented in the article are incredibly limited in scope, and it focuses almost entirely on implications for practice. The study did find that across all disciplines there was interest in disseminating theses and dissertations in electronic form using repositories, a practice which is growing at institutions across the U.S. (Shreeves and Cragin, 2008). It also identified some familiar barriers to IR participation, including technical confusion and copyright concerns.

In contrast to the small Rochester study, Lawal conducted a large, multi-institutional survey of science faculty that investigated self-archiving of explicitly pre-publication literature, articles that were intended for eventual publication (2002). The study focused on pre-print repositories, rather than IRs or DRs. The most common reason faculty gave for depositing their work in pre-print repositories was the desire for fast and widespread dissemination of their work. The study found dramatic differences by discipline; half of the physics and astronomy faculty surveyed share pre-prints, while none of the chemistry faculty do. The two main reasons faculty gave for not sharing pre-prints were that it was not relevant for their needs, and publisher policies prohibit it.

Factors That Influence Faculty Attitudes

As a group, studies on faculty attitudes seem to agree on the reasons that faculty do not self-archive, but offer conflicting results to explain why they do. Confusion about copyright and publisher policies, technical barriers, fears of plagiarism, and ignorance about IRs all arose repeatedly in these studies to explain why faculty are suspicious of self-archiving. What remains unclear are the roles that altruism and discipline have to play in encouraging self-archiving. Contradictory findings on these factors may mean that their effects are limited only to some faculty, or that the wide variation in methodology and phrasing used in these studies caused more variety in the results than exists in the population. Certainly the limited nature of many of these studies makes it difficult to draw broader conclusions from them. That there is consensus on barriers to self-archiving and disagreement regarding incentives and motivations suggests that careful work will be required to understand factors positively associated with sharing.

Faculty Self-Archiving Behavior

Impact of Publisher Policies

Given the studies on author attitudes suggesting that publisher policies are a prominent reason not to self-archive, studies examining the impact of those policies on actual behavior provide some valuable insight. Covey (2009) and Antelman (2006) employed similar methods to examine the impact publisher policies have on self-archiving behavior. Both pulled deposit data from the web without interacting directly with faculty, and both used the Sherpa/RoMEO database² on publisher self-archiving policies to compare behavior with policy. Covey conducted a case study using faculty websites at Carnegie Mellon University. Antelman selected journals from six social science disciplines, some that permitted self-archiving and some that did not, and searched for articles from those journals to determine whether a free online copy was available. Like Kim (2010), the two studies used broad definitions of self-archiving that included posting a link on a personal or department website along with deposit in an IR.

The parallels between the studies continue into the results sections: both found absolutely no connection between author self-archiving behavior and publisher policies. In other words, an author was just as likely to self-archive an article that had been published in a journal that prohibited self-archiving as she was to self-archive an article from a journal that permitted it. In Antelman's study, articles published in journals that prohibit self-archiving were *more* likely to be freely available online. The two studies also found that in contrast to publisher policies, discipline was a consistent predictor of self-archiving behavior. Covey and Antelman both infer that these behavior patterns indicate widespread ignorance of publisher policies on self-archiving, and the powerful influence of disciplinary norms.

Influence of Disciplines

A common finding throughout the literature has been that disciplinary norms influence

² <http://www.sherpa.ac.uk/romeo/>

faculty self-archiving practice (Kim 2011; Bell et al., 2005; Antelman, 2007; Covey, 2009). Two studies by Jingfeng Xia set out to test that result and found no relationship between disciplinary culture and self-archiving (2007; 2008). The first study pulled collection and deposit data from the IRs of seven universities in Australia across four disciplines – physics, chemistry, economics, and sociology – and produced weighted deposit rates based on the size of the departments (Xia, 2007). There were no clear patterns within a single discipline across multiple institutions. However, using the depositor field, Xia found that for some departments a single individual such as an administrative assistant or librarian made dozens of deposits on behalf of several authors, and that the presence of mediated support did help explain some of the variation across institutions.

Physics has one of the most robust disciplinary repositories on the internet, arXiv.org, and as a field is widely believed to have a strong culture of self-archiving (Kling et al., 2003). In his second study, Xia used the case of physics to investigate the theory that disciplinary culture, and particularly the presence of a robust disciplinary repository, has a positive effect on self-archiving practice and IR use (2008). Based on a published list of 51 Physics and Astronomy faculty at Southampton University in Australia, Xia attempted to identify all of the articles that were authored or co-authored by Southampton faculty, in both arXiv and the institutional repository. Though he reports deposit rates without applying any statistical analysis, his findings are intriguing: while many more faculty deposit in arXiv than in the IR, articles that are already in arXiv are much less likely to have been deposited in the IR. Participation in arXiv does not seem to predict participation in the IR. Xia suggests that once an article is available online in one place, faculty are less likely to archive it elsewhere (2008). This aligns with the finding at Cornell that the presence of a personal or departmental website where faculty could share their

work discourages them from using the IR (Davis and Connolly, 2007).

Influence of Mediated Deposit

Xia and Sun (2007) probed Xia's finding above that the presence of a librarian or staff person who managed deposits on behalf of faculty had a positive effect on IR participation (2007). Mediated deposit might have the potential to circumvent many of the barriers to self-archiving identified by research on author attitudes; it does not matter if the technical aspect of depositing is confusing or time consuming if someone else handles it on behalf of the researcher. This study used data from nine large IRs in the UK, Europe, and Australia that use the E-Prints software platform, which has a separate "Deposited By" field to identify the depositor as potentially distinct from the author. The contents of each IR were sampled across disciplines, though disciplinary results are not reported; the authors simply conclude that a large majority of the contents of all the IRs were not deposited by authors, but instead by a librarian, department secretary, or an automated process. This finding suggests that the very notion of "self-archiving" is a myth, and that while there is a practice of posting scholarly work online that is distinct from the formal publication process, it remains mediated by people other than the author herself (Xia and Sun, 2007). Such a result provides a possible explanation for those case studies that have had unexpected results with regard to disciplinary effects (Covey, 2009; Lawal, 2002; Lercher, 2008); perhaps those departments that displayed higher than expected deposit rates simply had a particularly engaged liaison librarian. At this point, however, such an explanation is purely speculative, and the study has limitations that mitigate against making such generalizations. It relied exclusively on publicly available data, sampled only large IRs that use Eprints, and made inferences based on the deposit data without qualitative evidence to support these conclusions.

Factors That Influence Faculty Behavior

Research on faculty self-archiving behavior has thus far focused on three factors: the impact of publisher policies, the influence of disciplines, and the influence of mediated deposit. While confusion about publisher policies was identified as a deterrent to self-archiving by research into faculty attitudes, the studies on faculty behavior suggest that confusion and ignorance are not necessarily barriers when it comes to self-archiving practice. Researchers who wish to post their articles online do so regardless of publisher policies (Covey, 2009; Antelman, 2007). Meanwhile, the influence of disciplines on faculty behavior remains an open question. Xia's findings that there is no relationship between discipline and self-archiving behavior contradict those of both Kim (2010) and Lawal (2002); clearly, more research is needed. The research reviewed here suggests that much of what is described as self-archiving behavior is in fact assisted by others. If this true more broadly, it may be the case that many of the barriers and incentives to self-archiving do not matter nearly as much as whether or not a given researcher has access to deposit support from her library or department. Again, further study is necessary to fully understand the role of mediated deposit in influencing self-archiving behavior.

Methodological Approaches

Most of the empirical research in this review suffers from at least one of two shortcomings: a lack of theoretical grounding, and a disregard for basic statistical principles. Contributing to this problem is that many of the studies included here were conducted by librarians, most of whom have neither doctoral degrees nor training in research methods beyond literature searching. Many academic librarians are required to publish peer-reviewed articles for tenure; this results in the publication of research studies written, reviewed, and edited by librarians without methodological expertise (Hoggan, 2003). Furthermore, many of the more

rigorous studies on open access and self-archiving, with larger sample sizes and more advanced statistical methods, have been released as reports or published in non-peer-reviewed venues such as D-Lib Magazine, and hence are not included in this review.

Almost all of the studies were exploratory in nature and did not test a clearly defined hypothesis. For the most part, researchers investigated faculty attitudes with surveys or interviews (Bell, Foster, and Gibbons, 2005; Gadd, Oppenheim, and Proberts, 2003; Kim, 2007; Kim 2010; Kim 2011; Lercher 2008; Watson, 2007), while those focusing on faculty behavior accessed deposit data directly from repositories and analyzed them quantitatively (Covey, 2009; Lawal 2002; Xia, 2008; Xia, 2007; Xia and Sun, 2007; Sale, 2006). The majority of the articles, both quantitative and qualitative, simply reported their results as unweighted percentages, and provided interpretations without regard for sample size, random sampling, response rates, or statistical significance. One study does not even report its sample size, or the methods used to obtain the sample (Bell, Foster, and Gibbons, 2005). As such, it is unclear how much we can rely on the accuracy or generalizability of this research, though to the extent that some of the studies reinforce each other and past findings, one could hope that they do contain some kernels of truth.

It must be noted that two researchers, Kim and Xia, have published multiple articles that defy this trend. Kim grounds her work in theory, uses statistical methods such as logistic regression and factor analysis, and tests her surveys in pilots before conducting large scale studies that rely on them. Xia explicitly sets out to test hypotheses, and carefully recounts methods used so that his studies may be reproduced by others. These basic hallmarks of good research would not be worth mentioning were they not unusual in the field as a whole.

Theoretical Limitations

Only two authors included in this literature review made any use of theory: Kim (2007;

2010) applied Socio-Technical Information Network (STIN) theory and social exchange theory in two of her articles, while Watson (2007) utilized a grounded theory approach to analyzing her interview data. Of these, the STIN model is the only one designed specifically to address the communication of knowledge and information in a networked context (Kling, McKim, and King, 2003). It was designed to apply to any form of electronic communication, but the set of heuristics that it offers appear to work well when applied to the question of scholarly self-archiving (Kim, 2007). Some of the most relevant heuristics include: “Identify a relevant population of system interactors; identify incentives; identify excluded actors and undesired interactions; identify existing communication forums; identify resource flows” (Kling, McKim, and King, 2003, p. 57). These heuristics could have provided some valuable context to many of the studies included here, and permitted a more systematic approach to investigating both the barriers and motivations for self-archiving behavior. In fact, the article that describes the STIN framework uses arXiv.org as an example for how to model an electronic scholarly communication forum (Kling, McKim, and King, 2003). One drawback of the STIN model is that it has not been further developed since 2003, and it does not appear to be widely used. Combined with the model’s focus on technical infrastructure, these limitations may cause the STIN model to become outdated over time, if it is not already.

The lack of theoretical grounding in the empirical research is particularly striking given the growing corpus of conceptual articles that apply a wide range of theories, both old and new, to the issue of open access. To be fair, online repositories and open access are only as old as the public internet, and scholars have been studying them for barely 15 years. Though relevant theories already exist for networked information (Benkler, 2003), there has not been much time to develop a robust theory for understanding open access. A few attempts at creating conceptual

frameworks for open access were published in the last four years (Pyati, 2007; Quinn, 2010; Herb 2010, Xia, 2011), making them too recent to have been applied to much of the research reviewed here. Looking ahead, more work is needed to develop and apply conceptual frameworks to the subject of open access broadly, and to the particulars of faculty attitudes and behaviors with regard to sharing their scholarly work online.

Implications for Research and Practice

Many questions remain about the motivations and barriers influencing authors' decisions to self-archive. In general, the field needs more large, multi-institution studies that investigate both faculty opinions about self-archiving and faculty behavior. Considering the conflicting results uncovered here, some broad, statistically rigorous investigations into the influence of disciplines on self-archiving behavior seem particularly warranted. The role of mediated deposit in increasing rates of self-archiving appears to be a promising area for future study, as well as one that could influence practice at libraries and other institutions that run repositories. If mediated deposit significantly raises rates of self-archiving, university libraries may choose to devote more resources to provide deposit support. Covey and Antelman's findings that faculty regularly disregard publisher copyright policies were based entirely on deposit data; it would be interesting to conduct qualitative studies to confirm just how much faculty do or do not know or care about publisher policies.

One question that has not been addressed in this review is the issue of open access mandates. In the last ten years, an increasing number of funders and institutions have implemented deposit mandates that require researchers to make their articles freely available online, regardless of where they are originally published (Suber, 2010). The purpose of mandates is to increase self-archiving behavior by making it a requirement of employment or funding (Sale,

2006). The NIH Public Access Policy, enacted in 2008, is the largest funder mandate in the United States and requires that all peer reviewed articles resulting from NIH grant funding be deposited in PubMed Central (Suber, 2010). While raw data released by the NIH indicates that deposit rates in PubMed Central have risen dramatically since the policy became law, there is currently no published research investigating the nuances of the policy's impact. Likewise, the impact of institutional mandates on faculty behavior in the United States remains unexamined.³ Much additional research is needed to understand more fully the potential benefits and downsides of open access mandates both at the institutional and funding levels.

Given multiple results suggesting that faculty prefer to use disciplinary repositories and personal or departmental websites to share their work online, it may also be prudent to take a hard look at whether institutional repositories are the best means to achieve the goal of open access to scholarly literature. What are the benefits of institutional repositories over other forms of self-archiving? For that matter, what are the benefits of self-archiving over other methods of making scholarly work freely available online? Is a system in which individual faculty deposit their articles into archives run by individual institutions the most efficient and effective way to improve access to scholarly literature and combat the high prices of for-profit publishers? These questions warrant prompt and careful study; the scholarly communications crisis that first inspired the open access movement shows no signs of abating on its own.

References

Antelman, K. (2006). Self-archiving practice and the influence of publisher policies in the social sciences. *Learned Publishing*, 19(2), 85-95.

³ One study examined the impact of institutional mandates on faculty self-archiving behavior at three universities in Australia and found that mandates resulted in 70-90% deposit rates, compared with 10-20% for institutions without a mandate (Sales, 2006).

- Bailey, C. W. (2006). What Is Open Access? In N. Jacobs (Ed.), *Open Access: Key Strategic, Technical and Economic Aspects* (pp. 13-26). Oxford: Chandos Publishing.
- Bell, S., Foster, N. F., & Gibbons, S. (2005). Reference librarians and the success of institutional repositories. *Reference Services Review*, 33(3), 283-290.
- Benkler, Y. (2003). Freedom in the Commons: Towards a Political Economy of Information. *Duke Law Journal*, 52, 1245-1276.
- Bergstrom, C. T., & Bergstrom, T. C. (2006). The Economics of Ecology Journals. *Frontiers in Ecology and the Environment*, 4(9), 488-495.
- Bergstrom, T. C. (2001). Free Labor for Costly Journals? *Journal of Economic Perspectives*, 15(4), 183-198. doi:10.1257/jep.15.4.183
- Emerson, R. M. (1976). Social Exchange Theory. *Annual Review of Sociology*, 2, 335-362. Retrieved from <http://www.jstor.org/stable/2946096>
- Gadd, E., Oppenheim, C., & Proberts, S. (2003). RoMEO Studies 2: How Academics Want to Protect their Open-Access Research Papers. *Journal of Information Science*, 29(5), 333-356. doi:10.1177/01655515030295002
- Herb, U. (2010). Sociological implications of scientific publishing: Open access, science, society, democracy, and the digital divide. *First Monday*, 15(2).
- Hoggan, D. B. (2003). Faculty Status for Librarians in Higher Education. *portal: Libraries and the Academy*, 3(3), 431-445. doi:10.1353/pla.2003.0060
- James-Gilboe, L. (2010). Raising the Library Profile to Fight Budget Challenges. *The Serials Librarian*, 59(3-4), 360-369. doi:10.1080/03615261003623112
- Joseph, H. (2008). A Question of Access — Evolving Policies and Practices. *Journal of Library Administration*, 48(1), 95-106.
- Kim, J. (2010). Faculty Self-Archiving: Motivations and Barriers. *Journal of the American Society for Information Science*, 61(9), 1909-1922. doi:10.1002/asi
- Kling, R., McKim, G., & King, A. (2003). A bit more to it: Scholarly Communication Forums as Socio-Technical Interaction Networks. *Journal of the American Society for Information Science and Technology*, 54(1), 47-67. doi:10.1002/asi.10154
- Lawal, I. (2002). Scholarly Communication: The Use and Non-Use of E- Print Archives for the Dissemination of Scientific Information. *Issues in Science and Technology Librarianship*, Fall.

- Lercher, A. (2008). A Survey of Attitudes About Digital Repositories Among Faculty at Louisiana State University at Baton Rouge. *The Journal of Academic Librarianship*, 34(5), 408-415. Elsevier B.V. doi:10.1016/j.acalib.2008.06.008
- National Institutes of Health. (2011). *NIH Deposit Statistics Through July 2011* [Excel spreadsheet]. Unpublished raw data.
- Park, J.-hong, & Qin, J. (2007). Exploring the Willingness of Scholars to Accept Open Access: A Grounded Theory Approach. *Journal of Scholarly Publishing*, 38(2), 55-84.
- Pyati, A. (2007). A critical theory of open access: Libraries and electronic publishing. *First Monday*, 12(10).
- Quinn, B. (2010). Reducing Psychological Resistance to Digital Repositories. *Information Technology and Libraries*, 67-76.
- Sale, A. (2006). The acquisition of open access research articles. *First Monday*, 11(10). Retrieved from <http://www.firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/1409/1327>
- Salo, D. (2010). The Battle of the Opens. *The Book of Trogoool*. Retrieved November 8, 2011, from http://scienceblogs.com/bookoftrogoool/2010/03/battle_of_the_opens.php
- Shreeves, S. L., & Cragin, M. H. (2008). Introduction: Institutional Repositories: Current State and Future. *Library Trends*, 57(2), 89-97. doi:10.1353/lib.0.0037
- Suber, P. (2010). Open Access Overview. Retrieved November 8, 2011, from <http://www.earlham.edu/~peters/fos/overview.htm>
- Watson, S. (2007). Authors' attitudes to, and awareness and use of, a university institutional repository. *Serials: The Journal for the Serials Community*, 20(3), 225-230. doi:10.1629/20225
- Wellen, R. (2004). Taking on Commercial Scholarly Journals: Reflections on the "Open Access" Movement. *Journal of Academic Ethics*, 2(1), 101-118. doi:10.1023/B:JAET.0000039010.14325.3d
- Willinsky, J. (2002). Copyright contradictions in scholarly publishing. *First Monday*, 7(11-4). Retrieved from <http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/1006/927>
- Xia, J. (2007). Assessment of Self-archiving in Institutional Repositories: Across Disciplines. *The Journal of Academic Librarianship*, 33(6), 647-654. Elsevier Inc. doi:10.1016/j.acalib.2007.09.020

Xia, J. (2008). A Comparison of Subject and Institutional Repositories in Self-archiving Practices. *The Journal of Academic Librarianship*, 34(6), 489-495. Elsevier B.V. doi:10.1016/j.acalib.2008.09.016

Xia, J. (2011). Constructing the structure underlying open access practices. *Journal of Information Science*, 37(3), 322-331. doi:10.1177/0165551511404868

Xia, J., & Sun, L. (2007). Assessment of Self-Archiving in Institutional Repositories: Depositorship and Full-Text Availability. *Serials Review*, 33(1), 14-21. doi:10.1016/j.serrev.2006.12.003