Self-archiving practice and the influence of publisher policies in the social sciences

Kristin Antelman
North Carolina State University

ABSTRACT: *Authors in different disciplines exhibit very different behaviours on the so-called 'green' road to open access, i.e. self-archiving. This study looks at the self-archiving behaviour of authors publishing in leading journals in six social science disciplines. It tests the hypothesis that authors are self-archiving according to the norms of their respective disciplines rather than following self-archiving policies of publishers, and that, as a result, they are self-archiving significant numbers of publisher PDF versions. It finds significant levels of self-archiving, as well as significant self-archiving of the publisher PDF version, in all the disciplines investigated. Publishers' self-archiving policies have no influence on author self-archiving practice.*

Kristin Antelman

Introduction

While repositories are transforming select disciplines, the more widespread impact of networked technologies has been to facilitate informal scholarly communication. Formal communication remains unchanged, at least for the time being; the peer-reviewed journal still constitutes the primary channel for official dissemination of scholarship. Where in the past scholars would have phoned a colleague to request a reprint, most of them will now email the colleague and receive an electronic copy of the article in reply. Some authors are choosing to take the next step in making their articles available by self-archiving a copy on a personal website or in a repository. There are several important consequences of this behaviour. One is that, by self-archiving, an author is contributing (knowingly or not) to the body of open access scholarship. Another is that, as these copies are distributed on the Internet, the reader is more likely to find versions of articles that may differ from the final published version.

Publishers and others have discussed the importance of understanding and controlling document versions from the perspective of scholarly communication.[1] At the heart of this issue is integrity, or trust. In our current system publishers facilitate peer review, provide a guaranty of authenticity and definitiveness, and brand an article. These quality filters are important both to authors and readers. When choosing where to submit a manuscript, authors attach the highest value to the journal's reputation,[2] and readers similarly rely on that context as affirming the quality of an article. This context is easy to assess with electronic journals, in which articles typically have the same look and feel as their print versions. But context is much more variable when looking at an article discovered through a search engine. The greatest context will be present with the

document version that looks exactly like the article retrieved from a publisher's website: the typeset and branded published article in PDF format.

While some publishers do not permit any self-archiving, most now permit authors to archive a version of their article, typically the author's final version. But because of its inherent substitutability for an article on the publisher's website, publishers who permit self-archiving typically do not grant authors permission to self-archive the publisher PDF version. If authors are self-archiving for scholarly communication, however, one would expect that they would archive the best and easiest version possible, which is the publisher PDF. There is therefore a tension between publishers' desire to control their content and some authors' desire to make it available in the best and easiest-to-use format. But authors may not even be aware that this tension exists. Studies have shown that authors are confused by, and generally uninterested in, their copyright agreements with publishers where they are most likely to learn about their self-archiving rights.[3] While several studies have explored various aspects of self-archiving behaviour, surveying authors on their practices and opinions or quantifying rates of open access, none have specifically focused on examining self-archiving of publisher PDFs.[4]

The study reported here looks at several aspects of self-archiving behaviour by authors publishing in leading journals in six social science disciplines. It tests the hypothesis that authors are self-archiving according to the norms of their respective disciplines rather than publisher self-archiving policies, and that, as a result, they are self-archiving significant numbers of publisher PDF versions.

Background

Versions

Mabe and Amin, and Guédon describe the two hats a scholar wears, that of researcher-as-author and that of author-as-reader.[5] The researcher-as-author cares relatively more about journals, while the author-as-reader cares relatively more about articles. As 75% of all authors search Google for articles, there are more readers of open access scholarship than there are authors practising self-archiving.[6] What factors into reader satisfaction with open access? Clearly, document findability and identifiability are fundamental. Guédon points out that we have a long way to go on findability, even for articles archived in disciplinary or institutional repositories (currently the minority).[7] Identifiability is a similarly confusing picture thanks to multiple versions.

Authors, publishers, open access advocates, and others have assigned labels to different versions of articles, but often these labels (e.g. preprint, postprint) do not describe the same objects. In identifying an article, a reader would take into account some combination of the document's content and who wrote it, attributes of the document's appearance such as formatting, notations about its provenance, and the context in which it is found. For example, one relatively well-defined, author-controlled version is the technical report or working paper preprint, usually so-labelled and located in the context of a series or collection. Institutional branding informs the reader that there has been some measure of quality control, while at the same time the reader knows from disciplinary acculturation that the text is likely to differ substantially from that of a final published article. Other preprints are much less well defined and could be a draft at any stage up to and including the final text. The SHERPA project website defines 'preprint' as 'the version of the paper before peer review' and the 'author postprint' as 'the version of the paper after peer-review with revisions having been made'.[8] This definition of preprint is not universally adopted, however, as the site notes: 'Another use of the term pre-print is for the finished article, reviewed and amended, ready and accepted for publication – but separate from the version that is type-set or formatted by the publisher. This use is more common amongst publishers.'[9]

In 2000, in a report entitled 'Defining and Certifying Electronic Publication in Science', a working group of publishers pointed to the importance of understanding and identifying article versions.[10] They opened

there are more readers of open access scholarship than there are authors practising self-archiving

the report with the statement, 'The peer-reviewed article will continue to play a crucial part in the certification, communication and recording of scientific research. However, in the electronic environment it represents one point on a potential continuum of communication.' In fact, in the context of open access, the peer-reviewed article represents at least two points on that continuum: the author's final version (the so-called author postprint), and the copy-edited, formatted, and branded publisher version. The author postprint is not only termed a preprint by some but it looks like one and can only definitively be identified as a postprint by comparing its text with the published article. While an author might make declarations of postprint status, such as 'accepted for publication in . . .', readers know at a glance that such a document is an author-controlled document and scholars trust publishers to assert this provenance, not individual authors. Without the contextual branding of a journal or pagination, such a document is not, according to the norms of most disciplines, citable.

Work on formal version control is still in the early stages. The report of the Research Councils of the UK on open access emphasizes the importance of distinguishing between pre- and postprints and states their intention to work with repository managers 'to develop a common and recognizable standard to ensure that the distinction between pre-prints and post-prints is clear to all users'.[11] In late 2005 JISC announced it would fund a scoping study on repository version identification, and NISO is also intending to explore the issue.[12] The NIH defines what they mean by postprint, and their policy for replacing it with the publisher's version if the publisher chooses, but they are not validating submissions as postprints.[13] In any event, such formal version control mechanisms are not likely to be adopted by authors posting documents to personal web pages.

Publisher self-archiving policies

Many, if not most, copyright agreements are silent and/or ambiguous about self-archiving rights and on the question of version. Even where the agreement addresses self-archiving, by using ambiguous terms to describe the versions (e.g. 'the work', 'the paper', 'the contribution'), the author does not have enough information to know what version he or she can self-archive. In addition, publishers require copyright assignment at different points in the review process depending on the journal, and cover different – or multiple – versions.[14] Ambiguity also exists in the more standardized self-archiving conditions in the SHERPA database. A survey of those conditions showed that, of 34 conceptually distinct conditions across all publishers, 12 could be associated with either preprints, postprints, or both, either because the restriction is unspecified or it makes sense for both (e.g. 'on author's or employer's website only').[15]

Very few authors' self-archiving behaviour is driven by what they find in SHERPA because few are aware of it.[16] More surprising, many authors are not aware of the terms of their copyright agreements and many believe they hold copyright to their own works.[17] At the same time, authors who know their rights or publisher requirements do not necessarily abide by them. Pinfield, and Swan and Brown found that there is clear evidence that a significant percentage of authors are self-archiving irrespective of the terms of their agreements, or self-archiving without pursuing clarification from publishers.[18] Lack of awareness of, or real interest in, the importance of these agreements is evidenced by how few authors attempt to modify them.[19]

A study of author behaviour in six social science disciplines

Methodology

Author self-archiving practice was surveyed across six social science disciplines: sociology, anthropology, economics, political science, geography, and psychology. Where possible (sociology, anthropology, and economics), journals with different author self-archiving policies were selected. The policies were 'nothing allowed' (SHERPA 'white' journals) and 'preprint and author postprint allowed' (SHERPA 'green' jour-

authors who know their rights or publisher requirements do not necessarily abide by them

nals). Publishers whose policies were not known to have changed since 2003, although they may have been codified since then, were selected. Eleven publishers are represented in the sample.[20] It is important to note that, despite what we know about publishers' policies and aggregate author behaviour, without asking individual authors it is impossible to know which rights they had, or believed they had, when they self-archived a given article, and whether they heeded or disregarded them.

Approximately 1,400 articles from 22 journals published between January 2003 and June 2004 were sampled. Google was searched to identify self-archived open access articles.[21] Because author behaviour, rather than overall rate of open access, is of interest, only those articles evidently posted by authors were counted. Open access copies found were coded by version: working paper, preprint, author postprint, or publisher PDF.[22] A separate sample of approximately 550 articles from the first half of 2005 was also taken to answer a related question, namely do authors self-archive at the same rate within six months of publication? That sample showed that the overall rate of self-archiving was virtually identical between the two time periods. Therefore, what follows reflects a combined data set of approxi-

without asking individual authors it is impossible to know which rights they had

mately 2,000 articles published between 2003 and 2005.

The author postprint is the version closest to the publisher's that most green publishers permit authors to self-archive, and is therefore a critical version from the perspective of open access. But, as Goodman points out, little is known about the actual prevalence of author postprints.[23] Acknowledging that it is impossible to identify definitively an author postprint without comparing texts, it remains of interest to approximate how many self-archived articles a reader would be likely to identify as author postprints. To understand how authors self-identify postprints, standard phrases were collected from title pages of documents examined in this study. The following designations were selected as likely identifying postprints: 'I would like to thank anonymous referees' (the most prevalent); 'published in/is forthcoming in/to appear in/accepted for publication in [journal name]'; or 'final version'. Using this admittedly approximate method to identify postprints, it was found that, within these six social sciences, the author postprint is overall quite rare (5% of all articles, constituting 16% of all self-archived articles). Pinfield has found that apparent preprints in arXiv were almost always in fact author postprints.[24] So to get a
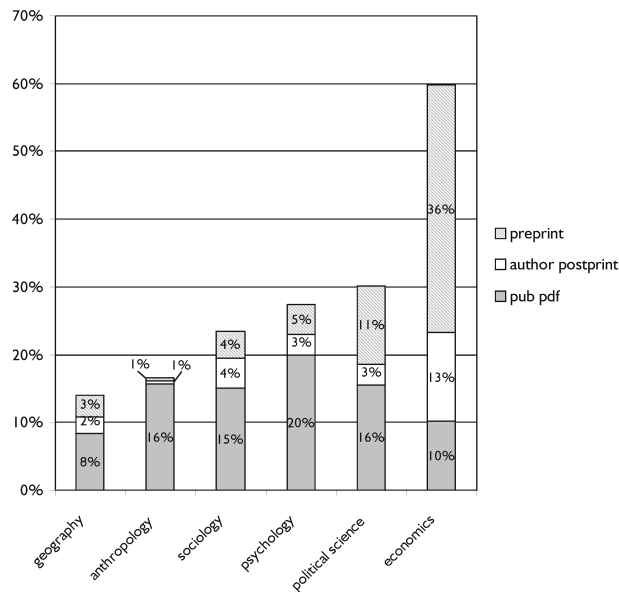

Figure 1 Self-archiving by discipline and article version.
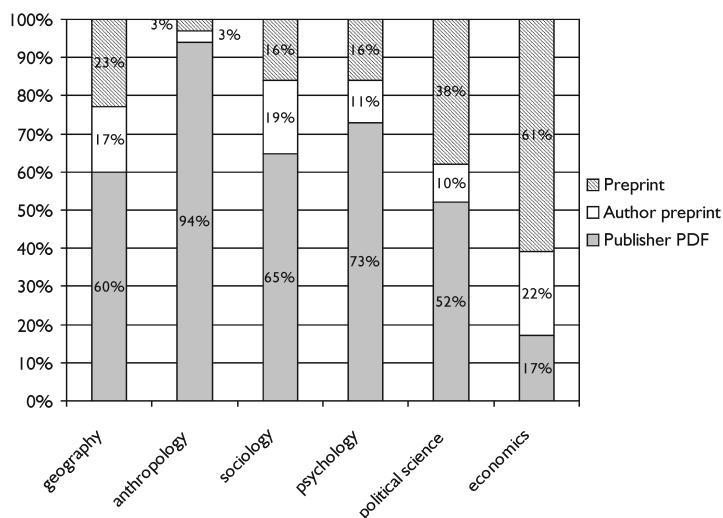
Figure 2 Versions of self-archived articles by discipline.

very rough sense of how many author postprints were missed using this approach, ten preprints were sampled across disciplines and compared with the publisher's version. Of those ten, five were indeed author postprints and five were preprints. So the actual number of author postprints in this sample is likely to be greater (and, therefore, preprints fewer).

Self-archiving rates by version

Self-archiving rates varied considerably across disciplines in the sample, from a low of 14% in geography to a high of 60% in economics (see Figure 1). It is important to note that the self-archiving rate in this sample cannot be generalized to the disciplines as a whole, since the sample contained disproportionately high-impact journals in each field, and there is evidence that, at least in biomedicine, authors publishing in prestigious journals archive at a greater rate.[25]

The version authors chose to self-archive varied significantly as well. Looking only at the self-archived articles (*n* = 575), authors overall posted approximately the same number of publisher PDFs (49%) as preprints/author postprints (51%). (See Figure 2.) This overall balance does not carry down to the individual disciplines, however: the percentage of self-archived articles that were preprints ranged from 3% in anthropology to

61% in economics; author postprints ranged from 3% in anthropology to 22% in economics; and publisher PDFs ranged from 17% in economics to 94% in anthropology. Interestingly, economics, with the highest overall self-archiving rate by a factor of two, has the lowest relative rate of self-archiving of the publisher PDF. Publisher PDFs made up between 52% and 74% of self-archived articles in sociology, political science, geography, and psychology. Excepting economics, the balance is 68% publisher PDF/32% pre- and author postprint.

The ease of working with the publisher PDF, for both authors and readers, surely contributes to its prevalence. Posting a publisher PDF is technically straightforward, particularly if the publisher provides the author with an electronic file of the version or final proof as part of the publication process. As Watkinson points out, in addition to ease of printing (appealing to readers), 'the other advantage of PDF was, and is, the simple built-in security'.[26] Posting an author postprint can be much more complicated, particularly when an author has multiple versions and, with charts or graphics, multiple files. Putting them all together into a single document some time after an article was submitted for publication can be a non-trivial process. That authors see these advantages of the PDF format apart from the branding advantage is evidenced by the fact

self-archiving rates varied considerably across disciplines

that over 90% of the author versions in this study were converted to PDF format before self-archiving.

Self-archiving rates and publisher policies

This study finds no relationship between publisher policy and self-archiving behaviour. In fact, the overall self-archiving rate for the white journals examined in this study is significantly *higher* than the self-archiving rate for the green journals (36% versus 27%, $\chi^2 = 11.62$, $P < 0.001$, $n = 6$ [white], $n = 16$ [green]).[27] Since white journals prohibit all self-archiving, clearly their policies are not influencing author behaviour. The same

absence of correlation can be seen in the rate of self-archiving of publisher PDFs. There is no significant difference when the six disciplines are looked at in combination: authors posted publisher PDFs of 15% of all articles from white journals and 14% from green journals. (See Figure 3.) Since more authors publishing in white journals self-archived, however, the percentage of self-archived articles represented by publisher PDFs is lower (42% for white journals versus 51% for green journals).

The correlation of publisher policy with self-archiving practice was examined within the three disciplines where journals with different policies were sampled (sociology, anthropology, and economics). (See Figure 4.) In economics, journal policy has little or no effect, with identical postprint and publisher PDF self-archiving rates for articles published in white and green journals and one-third more preprints in green journals. The possibility that policies have an effect in anthropology remains open. Authors of articles in white journals self-archive significantly fewer articles overall ($\chi^2 = 4.78$, $P < 0.05$). In sociology the picture is reversed, with significantly *more* overall self-archiving, and more self-archiving of publisher PDFs, of articles published in white journals (this could be explained in part by the fact that the flagship journal of the American Sociological Association, the *American Journal of*
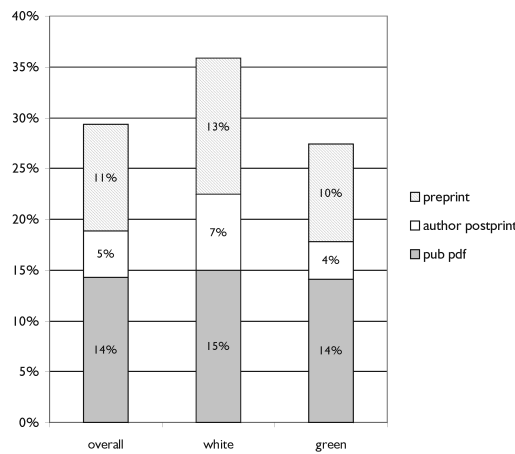

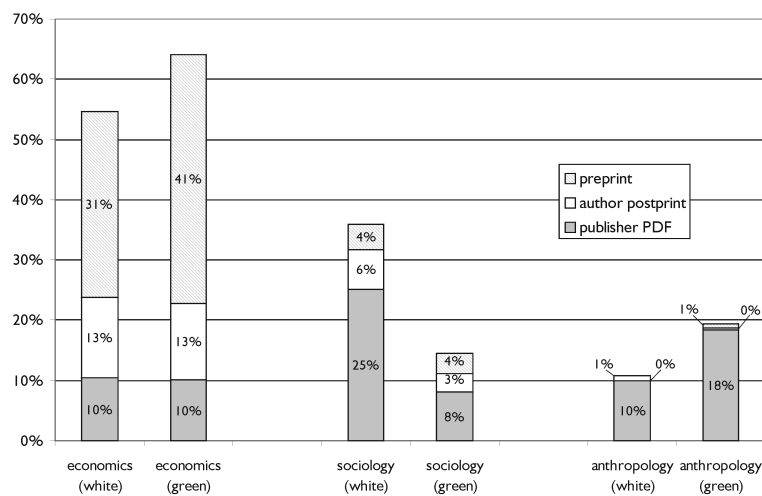**Figure 3 Self-archiving by journal policy and article version.**

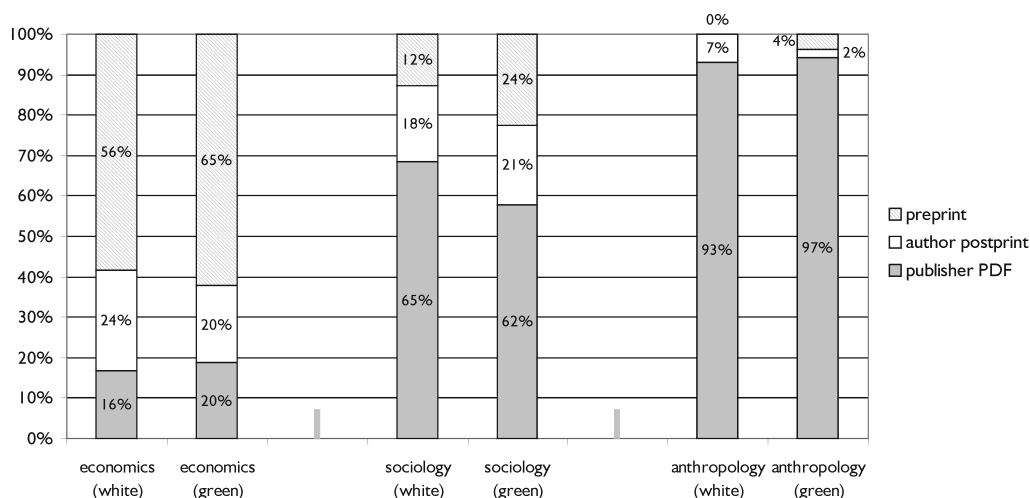**Figure 4 Self-archiving by discipline, publisher policy, and version.**

Figure 5 Self-archiving by discipline and publisher policy.

Table 1 Self-archiving rates by journal

	Self-archiving rate (%)				Overall self-archiving rate (%)
	Journal A	Journal B	Journal C	Journal D	
Economics	79 (g)	67 (w)	56 (g)	41 (w)	60
Sociology	48 (w)	26 (w)	12 (g)	11 (g)	23
Anthropology	24 (g)	16 (g)	16 (w)	5 (w)	17

w, white; g, green.

Sociology, is a white journal and has a self-archiving rate of 48%).

A disciplinary norm for sharing versions appears to be consistent across publisher policies. (See Figure 5.) The publisher PDF as a percentage of all self-archiving is virtually the same in each discipline. At the least, one can conclude that distribution of preprints is important in economics, less so in sociology, and much less so in anthropology.

Underscoring the tentative nature of these differences between disciplines, it was observed that there is more variation between the journals within the three disciplines than there is between the disciplines. (See Table 1.) Significant variation is also seen between journals published by the same publisher. For instance, the four journals published by the University of Chicago Press had overall self-archiving rates between 16% and 67% and ranged between 16% and 91% in publisher PDF as a percentage of all self-archiving. Author self-archiving rates in

the two University of Chicago Press titles in economics were also significantly different (67% and 41%).

To look at the question of whether self-archiving practices were consistent within the same journal over time, articles from 2003 were compared with articles from 2004–5 for each journal. The similar overall self-archiving rates (28% in 2003, 29% in 2004–5), and relatively consistent self-archiving rates at the discipline level, concealed great variability at the journal level. Seven journals had close to or the same self-archiving rates between the time periods, 11 increased, and seven decreased, ranging from a 115% increase to 100% decrease in self-archiving. This fluctuation reflects small sample sizes at the individual journal level (averaging 46 articles for each time period). Given this great variability at the journal level, it is intriguing that the rates remained relatively constant at the discipline level and overall.

distribution of preprints is important in economics, less so in sociology, and much less so in anthropology

Discussion

This study's findings that the self-archiving profiles of each discipline examined are very different, and that there is an evident lack of influence of publisher policies across the board, both point to the fact that disciplinary norms for scholarly communication are driving author self-archiving behaviour. There are several conceptual models that can help frame these findings. Whitley proposed that differences between disciplines can be characterized in terms of the degree of mutual dependence between researchers and the degree of task uncertainty in defining shared problems, goals and procedures.[28] Disciplines with high mutual dependence and low task uncertainty move very quickly to answer problems that have agreed-upon importance, and they place a high value on establishing priority (high-energy physics, the first field to self-archive and establish a disciplinary repository, is the prime example). As he points out, the humanities and social sciences tend to be characterized by a low degree of mutual dependence and high task uncertainty.[29] Researchers in these disciplines address a wide range of research questions and lack shared problem definitions, theoretical goals, and work procedures.[30] As a result, they tend to rely less on exchange of pre-published research.

Becher proposes an analogous model for describing the social dimensions of intellectual fields using two properties, convergent/divergent and urban/rural.[31] In divergent disciplines, just as in disciplines with low mutual dependence, there is a less stable élite with less intellectual control over the field than in convergent disciplines. While economics is characterized as convergent, sociology and geography are characterized as divergent.[32] While Becher does not address them specifically, political science, anthropology, and psychology would also be more divergent than convergent. The six social sciences examined in this study, with the exception of much of economics, are also considered rural disciplines, characterized by low people-to-problem ratios and research that addresses a wide range of problems without agreed-upon priority or importance. Rural disciplines rely less on sharing preprints, as rapid dissemination of results and establishment of priority is of lesser priority than in urban disciplines.

This study confirms the profile of the rural, low mutual dependence/high task uncertainty discipline in finding relatively few self-archived preprints in anthropology, geography, sociology, and psychology. This was less the case in political science (11% preprints) and not the case at all in economics (36% preprints). The relatively high number of working papers in economics (more than one-third of all preprints) shows that economics has characteristics of a discipline with mutual dependence in engaging in informal sharing of research results in symposia and working paper series. These disciplines, with economics and political science again being most anomalous, exhibit another characteristic of rural disciplines in most frequently self-archiving postprints.

Another characteristic of rural fields is porous disciplinary boundaries. In economics and political science in particular, and to a lesser extent sociology, psychology, and geography, authors have – and desire to have – readers from related disciplines and even outside academia, for instance in government. This pursuit of a broad audience could be an incentive for some authors to self-archive. Since fields that are most concerned with establishing priority have the shortest publication delays,[33] the presence of self-archiving in the social sciences may not be due so much to authors' desire to establish priority, but to the desire of a significant minority to use readily available technologies to mitigate the detrimental effects of particularly long publication delays. At least one study has shown authors are quite concerned with such delays.[34] On the other hand, the relative absence of disciplinary repositories in the social sciences indicates that informal publication on personal or departmental websites, however incompletely indexed such websites are, seems to suffice to meet authors' needs.

Conclusions and further study

This study finds that social scientists are self-archiving at a significant rate and that publisher self-archiving policies do not influ-

there is an evident lack of influence of publisher policies across the board

ence their self-archiving practices. The likelihood that an author will self-archive at all varies between disciplines but remains fairly constant between different publisher policies ('green' and 'white') in the three disciplines where journals with different policies were examined. Similarly, the rate at which authors self-archive the publisher PDF version, a practice that none of the surveyed journals permits, varies significantly between disciplines but not by journal policy.

Many authors as well as publishers consider posting an article to a website or repository to be a form of publication. In King's framework, self-archiving would be considered 'weak' publishing, while the published, indexed version in the journal would be 'strong' publishing.[35] From the perspective of the author-as-reader searching on Google for desired articles, the strongest publications are those that are both discoverable and useful, i.e. have accessible full text and are a citable version. A publisher copy that the reader does not have access to, while discoverable, is not accessible and so not as useful as an open access copy. Similarly, an open access copy located on a website is not as discoverable as it could be if it were in a repository or a journal. On the utility side, a preprint or author postprint is less useful than the publisher's version because, while it supports knowledge discovery, it cannot be cited. The self-archived publisher PDF, while not as discoverable as the publisher's copy, is just as useful because it bundles knowledge discovery with certification and citability. The publisher PDF version allows the author to transfer the value of the publisher's copy to one he controls, and allows readers to apply what they already know about identifying scholarly publications. Thus, it is not surprising that most authors who choose to self-archive a postprint, self-archive the publisher PDF version, the version that is both easiest to post and that they know will provide readers with the information they want.

As 'rural' disciplines, and therefore with a less strong sense of collective identity, the social sciences (excepting economics) self-archive at a moderate rate, nothing like what is seen in more 'urban' fields such as physics, mathematics, and computer science that have developed disciplinary repositories to meet the demands for rapid communication. A corollary of this is that which journals one selects to examine is particularly consequential when seeking to understand behaviour at a discipline level. As Guédon notes, journals in the social sciences and humanities 'often tend to incarnate a theoretical position or even a particular group, rather than a segment of knowledge.'[36] The great variability found between journals within disciplines in this study underscores that fields with high task uncertainty tend to create journals that occupy distinct intellectual niches. As Fry and others have demonstrated, intellectual fields within a single discipline can vary to a great extent, and a given intellectual field may have more in common with an intellectual field in another discipline than its own parent discipline.[37] In fact, both Whitley and Becher prefer to make comparisons at the level of intellectual fields (or specialisms) rather than disciplines. Future studies selecting intellectual fields as the unit of analysis rather than disciplines will be better positioned to identify and interpret differences in self-archiving practice. Further light could be shed on the question of the influence of publisher policy on self-archiving behaviour if authors were the unit of analysis, and the actual status of their copyright agreement, as well as their knowledge and stated motivation, were correlated with their self-archiving behaviour at the article level.

While it is clear that by self-archiving an author is contributing to the body of open access scholarship, his action may or may not reflect self-aware endorsement of open access principles. This study shows that, for many reasons, one should be careful about characterizing a given rate of self-archiving as a rate of 'adoption' of open access, and authors as 'complying with' or 'violating' publishers' policies. Self-archiving one's own research may be simply a logical extension of existing scholarly communication practices into the digital realm. Just as it is authors and not publishers who self-archive, it is discipline-based norms and practices that shape self-archiving behaviour, not the terms of copyright transfer agreements.

most authors who choose to self-archive a postprint, self-archive the publisher PDF

References

1. Watkinson, A. Securing authenticity of scholarly paternity and integrity, Book Industry Communication, 2003. (http://www.bic.org.uk/securing%20 authenticity.pdf). Frankel, M. S., *et al.* Defining and certifying electronic publication in science: a proposal to the International Association of STM Publishers. *Learned Publishing*, 2000:13(4), 251–8. URL: www.ingentaselect.com/rpsv/catchword/alpsp/09531513/v13n4/s8/p251. Lynch, C. When documents deceive: trust and provenance as new factors for information retrieval in a tangled web. *Journal of the American Society for Information Science and Technology*, 2001:52(1), 12–17.

2. Rowlands, I. and Nicholas, D. New journal publishing models: an international survey of senior researchers. A CIBER report for the Publishers Association and the International Association of STM Publishers, 2005, p. 17 (http://www.slais.ucl.ac.uk/papers/dni-20050925.pdf).

3. Swan, A. and Brown, S. Open access self-archiving: an author study. Key Perspectives Limited, 2005, p. 56 (http://eprints.ecs.soton.ac.uk/11006/).

4. Swan and Brown take a somewhat different approach and find that 49% of survey respondents had self-archived or published in an open access journal in the previous three years (Open access self-archiving, 30). Harnad, S. and Brody, T. Comparing the impact of open access (OA) vs. non-OA articles in the same journals. *D-Lib Magazine*, 2004: June. (http://www.dlib.org/dlib/june04/harnad/06harnad.html). Kurtz, M. J., Eichhorn, G., Accomazzi, A., Grant, C., Demleitner, M., and Murray, S. S. Worldwide use and impact of the NASA Astrophysics Data System digital library, *Journal of the American Society for Information Science and Technology*, 2005:56(1), 36–45. Antelman, K. Do open access articles have a greater research impact? *College & Research Libraries*, 2004:65(5), 372–82. Wren, J.D. Open access and openly accessible: a study of scientific publications shared via the internet. *BMJ*, 2005:330, 1–5. Wren's study did focus on PDFs, but the documents found were not necessarily self-archived by authors.

5. Mabe, M. A. and Amin, M. Dr Jekyll and Dr Hyde: author–reader asymmetries in scholarly publishing. *Aslib Proceedings*, 2002:54(3). Guédon, J. C. The 'green' and 'gold' roads to open access: the case for mixing and matching. *Serials Review*, 2004:30(4), 316.

6. Swan and Brown (Open access self-archiving, 25) found that 72% of researchers surveyed reported searching for articles in Google. Gadd *et al.* found that 75% of authors who had not self-archived (and 95% of those who had) had used others' self-archived papers. Gadd, E., Oppenheim, C., and Probets, S. RoMEO Studies 3: how academics expect to use open-access research papers. *Journal of Librarianship and Information Science*, 2003:35(3), 7.

7. Guédon, J. C. The 'green' and 'gold' roads to open access: the case for mixing and matching. *Serials Review*, 2004:30(4), 316.

8. Definitions and terms: pre-print and post-print, SHERPA Publisher Copyright Policies & Self-archiving (http://www.sherpa.ac.uk/romeoinfo.html).

9. *Ibid.*

10. Frankel, M. S., Elliott, R., Blume, M., Bourgois, J. M., Hugenholtz, B., Lindquist, M. G., Morris, S., and Sandewell, E. Defining and certifying electronic publication in science. *Learned Publishing*, 2000:13(4), 251. www.ingentaselect.com/rpsv/catchword/alpsp/09531513/v13n4/s8/p251

11. RCUK Position statement on access to research outputs, Research Councils UK, 7–8 (http://www.rcuk.ac.uk/access).

12. JISC ITT: Scoping study on repository version identification. (http://www.jisc.ac.uk/index.cfm?name=funding_versionidentification). NISO/ALPSP working group on versions of journal articles. (http://www.niso.org/committees/Journal_versioning/JournalVer_comm.html).

13. Policy on enhancing public access to archived publications resulting from NIH-funded research. US Department of Health and Human Services National Institutes of Health (http://grants.nih.gov/grants/guide/notice-files/NOT-OD-05-022.html).

14. Gadd *et al.* found that 69% of publishers asked for copyright transfer prior to refereeing the paper. Gadd, E., Oppenheim, C., Probets, S. RoMEO Studies 4: an analysis of journal publishers' copyright agreements. *Learned Publishing*, 2003:16(4), 293. www.ingentaselect.com/rpsv/catchword/alpsp/09531513/v16n4/s9/p293

15. This survey of the SHERPA database was conducted on 13 June 2005. It is also noted that a quantitative view of restrictions and conditions does not reflect the true impact on authors, because a publisher-level view treats small and large publishers equally. Also, not all conditions are equivalent in their burden on the author: Pinfield found that authors follow reasonable ones that are also in their own interest, e.g. citing the published version on a postprint, but not unreasonable ones, e.g. copying specified clauses from the copyright agreement. Pinfield, S. How do physicists use an e-print archive? Implications for institutional e-print services. *D-Lib Magazine*, 2001: Dec, 5. (http://www.dlib.org/dlib/december01/pinfield/12pinfield.html).

16. Swan and Brown (Open access self-archiving, 48) report that only 10% of authors are currently aware of SHERPA.

17. JISC Disciplinary Differences Report, Sparks, S., Rightscom Ltd, 2005, 49. (http://www.jisc.ac.uk/uploaded_documents/Disciplinary%20Differences%20and%20Needs.doc). Swan and Brown (Open access self-archiving, 56) report 35% believe they hold such rights. Gadd and Oppenheim found 61% of authors surveyed believed they held copyright. Gadd, E. and Oppenheim, C. RoMEO Studies 1: the impact of copyright ownership on academic author self-archiving. *Journal of Documentation*, 2003:59(3), 256.

18. Gadd *et al.* reported 12% of respondents to their survey would ignore publisher policies forbidding them from making articles freely available online. RoMEO 4, 13. Swan and Brown (Open access self-archiving, p. 56) reported that 84% of all authors did not ask for permission and that one-third did not know if permission was required. Pinfield randomly

sampled ten articles in arXiv that were published by publishers forbidding posting of the postprint and found that all were the same as published version in content (How do physicists use an e-print archive?, pp. 5–6).

19. JISC Disciplinary Differences Report, 49. Swan and Brown, Open access self-archiving, p. 56. A University of California faculty survey reported by John Ober indicated that 18% of respondents have at one time modified the terms of a copyright agreement or contract. Ober, J. Postprint Repository Services: Context and Feasibility at the University of California, Final Draft, 2005, 23 (http:// osc. universityofcalifornia.edu/responses/materials/UC_po stprintstudy_final.pdf). Gadd *et al.* reported that only 3% of authors insisted on retaining copyright, RoMEO Studies 1, 259.

20 'White' journals surveyed were: *American Journal of Sociology, Current Anthropology, Economic Development and Cultural Change,* and *Journal of Political Economy* (all University of Chicago Press); *American Sociological Review* (American Sociological Association); *American Ethnologist* (American Anthropological Association). 'Green' journals surveyed were: *Journal of Social Issues, Social Problems* (University of California Press), *Annals of the Association of American Geographers, Psychological Science, Econometrica,* and *American Journal of Political Science* (all Blackwell), *Political Behavior* and *Theory and Society* (both Kluwer), *American Journal of Physical Anthropology* (Wiley), *Journal of Human Evolution, Political Geography,* and *Journal of Econometrics* (all Elsevier), *Journal of Conflict Resolution* and *Comparative Political Studies* (both Sage), *Progress in Human Geography* (Arnold), and *Journal of Applied Psychology* (American Psychological Association). Data were collected between March and July 2005.

21. Only articles labeled by journals as 'research' or 'original' were selected. Typically, full titles were searched as phrases in Google. If there were no results, or there was ambiguity between documents with the same title, the author's last name was added to the query. Since Google does not consistently index parenthetical phrases or certain characters (e.g., ?, a), where present those were omitted from the queries.

22. When multiple self-archived versions were found, the "highest" (i.e., closest to the final publisher version) was coded as the self-archived version.

23. Goodman, D. The criteria for open access. *Serials Review,* 2004:30(4), 260.

24. Pinfield, S. How do physicists use an e-print archive? Implications for institutional e-print services, *D-Lib Magazine,* 2001: Dec, 5. (http://www.dlib.org/dlib/ december01/pinfield/12pinfield.html).

25. These numbers do not represent an accurate estimation of open access in these disciplines for several other reasons: non-Google search engines were not used, unindexed author homepages were not sought out, and only author-archived copies were counted.

26. Wren, J. D. Open access and openly accessible: a study of scientific publications shared via the internet. *BMJ,* 2005:330, 1-5.

27. Watkinson, A. Securing authenticity of scholarly paternity and integrity, Book Industry Communication, 2003. (http://www.bic.org.uk/securing%20 authenticity.pdf).

28. Whitley, R. *The Intellectual and Social Organization of the Sciences,* New York, Oxford University Press, 2000, chapter 5.

29. Ibid., 127.

30. Ibid., 119–20.

31. Becher, T. and Trowler, P. R. *Academic Tribes and Territories,* The Society for Research into Higher Education,& Open University Press, 2001, chapter 9.

32. Ibid., 187–88.

33. Ibid., 189.

34. 'What Authors Want': the ALPSP research study on the motivations and concerns of contributors to learned journals. *Learned Publishing,* 1999:12 (3), 171. www.ingentaselect.com/rpsv/catchword/alpsp/095315 13/v12n3/s1/p170

35. King, R., Spector, L., and McKim, G. The guild model. *Journal of Electronic Publishing* 2002:8(1). (http://www.press.umich.edu/jep/08-01/kling.html).

36. Guédon, J. C. In Oldenburg's long shadow: librarians, research scientists, publishers, and the control of scientific publishing. In ARL (Association of Research Libraries) Proceedings, 138. May 2001 Membership Meeting, 22. (http://www.arl.org/arl/proceedings/138/ guedon.html).

37. Fry, J. Scholarly research and information practices: a domain analytic approach. *Information Processing & Management* 2006:42 (available online Nov 2004), 299–316. Becher and Trowler.

Kristin Antelman
Associate Director for Information Technology
NCSU Libraries
Box 7111
Raleigh, NC 27696-7111, USA
Email: kristin_antelman@ncsu.edu