

Chapter 3 – Researchers and repositories

Introduction

The previous chapter described how open access offers solutions to many of the challenges facing the scholarly communication system. This chapter explores the idea that understanding the communication behaviours of the individual scholar is central to any large scale uptake of a new scholarly communication system, and that research into the field of repository use has, to date, been restricted to snapshots of the scholarly community via large internet surveys.

The chapter begins with a description of the two ways of achieving open access dissemination: publishing in open access journals and depositing articles in digital repositories. It then describes how open access will affect different members of the scholarly community. The chapter will then explore the question of who is sharing their work in this manner to help illuminate the question of why many scholars are choosing not to do this.

Achieving open access – the gold road

If, as Chapter 2 has demonstrated, open access offers a wider and more inclusive dissemination of information than the current scholarly communication system, the next question is: how is open access achieved? There are generally two ways, described as the ‘gold’ and ‘green’ roads to open access, referring to open access journals and digital repositories respectively (Harnad et al., 2004b). For reasons explained below, this research will focus more closely on repositories than on open access journals. However, to ensure a complete picture of the open access dissemination landscape is provided in this work, a brief discussion of the open access journal option is provided here.

As stated in Chapter 1, open access generally refers to the dissemination of research in a way that is freely available to any interested reader with an internet connection. The ‘gold road’ to open access dissemination refers to open access journals, broadly defined as journals that do not charge a subscription fee. Wilinksy (2006) offers ten ‘flavours’ of open access, several of which are ‘gold’. In some cases the journal operates without cost to the reader or the author, an example of ‘Cooperative’ type of open access journals, in other cases subsidies from scholarly societies or elsewhere provide the financial support for

‘Subsidised’ open access, and there is also ‘Dual-mode’ where the subscriptions for print editions also cover the cost of open access online.

A fourth gold flavour of open access Willinsky describes is ‘Author fee’, which is a business model that emerged to allow commercial publishers (and scholarly associations who rely on membership in exchange for journal publication) to publish open access journals and also continue in business. This is achieved by charging an up-front fee to cover the costs of reviewing and editing the paper, often referred to as the ‘author-pays’ model. This term is slightly misleading, as it is usually the author’s institution or funding body that covers the cost, so the term ‘pay on acceptance’ is more accurate. There are institutional models under this system, such as BioMed Central where the institution pays a membership for the researchers to publish in the open access journals. It should be noted, however, “OA journals that charge processing fees usually waive them in cases of economic hardship” (Suber, 2007b). Some examples of policies of fee waivers are (BioMed Central, 2008; Nucleic Acids Research, 2008; PLoS Public Library of Science, 2008).

There are potential issues with this system. For example the sustainability of a pay on acceptance model is particularly problematic for journals with high rejection rates (DeAngelis & Musacchio, 2004). The system that has authors paying for publication either directly or through sponsorship from institutions or interested third parties means that, “science will either have a less effective filter, or will require the introduction of new post-publication filtering mechanisms” (Crawford, 2003). The system is biased towards the author’s goal of publication rather than the reader’s goal of effective filtering. There is also a potential conflict of interest in the refereeing process, where “the incentives for editors in an author-pays model will be to publish more papers not fewer” (Horton, 2003). This issue has also been raised in surveys of authors’ attitudes: “Feedback has also highlighted the essential requirement that author charging and editorial decisions be completely separated and seen to be separated. Authors should be reassured that their ability to pay publication charges will not influence editorial decisions in any way” (Richardson & Saxby, 2004).

One method publishers have developed to introduce the idea of authors paying to have their work freely accessible is to develop a hybrid model for their journals. In this model, a journal offers authors the opportunity to publish their work in open access for a fee, and (in theory) reduces the subscription cost in proportion to the number of articles that are published in this fashion. If the author chooses not to pay a fee their work is only accessible to subscribers. Some of the journals offering a hybrid option have ‘anticipated’ the subscription cost of the journal will be reduced according to the number of open

access articles that appear in the issues (Suber, 2006). This is happening with at least a few publishers, with Oxford Journals announcing:

In 2008, the average increase across all Oxford Journals titles is 6.9%. For the 28 Oxford Open titles with open access uptake in 2006 (the last full calendar year on which we could calculate), the average 2008 online-only price increase is just 1.7%. This is due to adjustments to the online-only prices of these journals which reflect increases in the percentage of open access content published between 2005 and 2006 (Richardson, 2008).

There has been a recent surge in hybrid models, with the proportion of publishers offering this to authors growing from 9% in 2005 to 30% in 2008 (ALPSP, 2008). Of hybrid journals available to researchers in 2007, 12 were published by for-profit publishers, and 65 by societies. Only two hybrid journals were in the social sciences with the remainder in the STM fields. "Hybrid journals are clearly most common in the fields in which most authors have research grants from which they might be able to pay the journal's publication fee" (Suber & Sutton, 2007). There is little available evidence of the level of uptake of this option by authors, although apparently 18% of authors publishing in journals published by the American Physiological Society have done so (Biello, 2007). A 2008 study showed that "the take-up of the author pays open access option is exceedingly low" (ALPSP, 2008).

Some institutions have programs to pay for publication charges in open access journals, for example the University of Nottingham (2007) has established an Open Access Publications Fund and the Library of the University of Amsterdam (2007) has also offered an open access fund since January 2007. The University of California Berkeley launched a program in January 2008. Established programs have not had a high uptake, the University of North Carolina-Chapel Hill launched a program in March 2005 but only 14 grants had been allocated by May 2008. The University of Wisconsin-Madison program offering authors half of OA fees has allocated only 16 awards since its inception in 2006 (Adams, 2008).

A potential drawback of the hybrid model is the 'tipping point' when subscribers decide that enough of the journal is online to warrant canceling the subscription (Prosser, 2003). This is probably more of an issue with individual subscribers, as libraries tend to hold their subscriptions in bundles, which would make individual cancellations difficult. In fact the hybrid model initially increases costs to the institutions (universities, funders and governments) because they will be, at least for a while, paying for both kinds of publication (Suber, 2005b).

Open Journal Systems

Open access advocates have developed tools to help people wanting to start up an open access journal. Open Journal Systems (OJS)ⁱ is open source software that has been developed by the Public Knowledge Project at the University of British Columbia. It began as a proof of concept to test the degree to which an open source and easily configurable piece of software can reduce the cost of running a journal by moving the process online, not only in the publishing and distribution of the journal, but in its actual day to day management (Willinsky, 2006). OJS sets up a website for the journal which acts as an editorial office for editors, reviewers, authors, copyeditors and others. The program labels, files and tracks all submissions.

As one example of how OJS works, the National Library of Australia has adopted OJS and is successfully managing an online open access journal publishing service. The library began by migrating the journal from the Association for the Study of Australian Literature, JASAL, to an online format, using OJS to assist with every stage of the refereed publishing process, from submissions through to online publishing and indexing. It is now increasing its stable of open access journals (Graham, 2006).

Achieving open access – the green road

The alternative to open access journals, the ‘green’ method of open access dissemination, is achieved by making the author’s versions of articles available online. This can be through an author’s own website, although generally digital repositories are considered to be more ‘robust’ because individual websites are likely to die with the author or be lost if the author changes employment. Digital repositories are also searchable due to a requirement that they comply with the Open Access Initiative (OAI) Protocolⁱⁱ which requires interoperable standards for searching of repositories. The term ‘self-archiving’, while widely used in open access discussions, is inaccurate because the person who is doing the ‘archiving’ has no control over the long-term life of the item. A more accurate term, and one which will be used throughout this thesis, is ‘self-deposit’. It is important to note that depositing an article in a repository does not, in itself, constitute publishing.

The word repository can mean many things. This thesis is concerned with digital repositories which are: “usually considered to mean a place where one deposits ... objects such as peer-reviewed research manuscripts for the purpose of providing open access. The ... object needs to be captured at creation time otherwise it is soon lost” (Sale, 2005a). A simpler definition includes the descriptors: cumulative, perpetual, open and interoperable (Ware, 2004a). Incorporating all these definitions, repositories provide the

means for the alternative, 'green' road to open access, when an author can self deposit a pre-print, the digital text of a peer-reviewed research article before refereeing or post-print, a copy of the refereed article, in a digital repository.

Repositories are developed with different purposes in mind. This thesis is predominantly concerned with institutional and subject-based repositories. While this author has previously argued that "the term institutional repository is often narrowly applied to repositories run by universities and other research-oriented employers. It is more appropriately used as a broad term to encompass both those and the repositories that have been run by learned associations and other scholarly communities since as early as 1991" (Clarke & Kingsley, 2008), the discussion in this thesis requires a delineation between different types of digital repositories.

A functional definition of an institutional repository adopted in this thesis is: "A set of services that a university offers to the members of its community for the management and dissemination of digital materials created by the institution and its community members" (Lynch, 2003). An alternative type of repository is a subject-based repository. This term is used in the literature but is not the optimum word to explain this type of repository because the term generally refers to repositories that not only belong to a scholarly community, but also a discipline or research domain.

Despite the large amount of material written about digital repositories there is still a small proportion in existence in institutions world-wide. As of September 2008, 1145 repositories were listed on the Registry of Open Access Repositories (ROAR).ⁱⁱⁱ Of those, 611 are 'Research Institutional or Departmental'. The World List of Universities^{iv} on the same day lists 8136 universities in 198 countries. A simple calculation indicates that only 7.5% of universities worldwide have a repository with the correlating result that 92.5% do not. However, Australia is in a unique position worldwide because as of September 2008, 32 of the 39 universities have active repositories, of which 31 are open access. A further three universities are planning to launch their repositories later in 2008, two have plans to launch in 2009, and the closed access repository is planning to open access to items in 2009. This means 37 Australian universities should have active repositories by the end of 2009 (Kennan & Kingsley, 2009).

For those who argue that a digital repository's main function should be to provide free access to the literature, the breakdown of the content in repositories must be disappointing. A survey of 45 repositories undertaken in 2003 had an average number of documents per archive of 1256. The breakdown of these was 22% e-prints, including pre-prints and post-prints. Theses and dissertations made up 20%, with the remaining 58%

made up of other documents including grey literature, such as technical reports, and working papers (Ware, 2004b, p. 25). Despite the large number of Australian repositories, looking at the number of items in Australian repositories listed on the OpenDOAR^v website demonstrates that initial uptake of the repositories has been slow in that country. A more recent survey has shown that of the repositories in Australian universities, three have more than 5,000 open access items, the remainder have 1000 or fewer items (Kennan & Kingsley, 2009).

Clearly there is a barrier to scholars depositing their work in institutional repositories. Generally, subject-based repositories are enjoying a greater level of self-depositing than institutional repositories. As a demonstration of this preference, it is instructive to look at the participation levels of three subject-based repositories¹. According to its site, arXiv, at the time of writing, offers: “open access to 451,387 e-prints in Physics, Mathematics, Computer Science, Quantitative Biology and Statistics”. RePEc (Research Papers in Economics)^{vi}, holds 222,000 working papers, 316,000 journal articles, 1,500 software components and numerous listings from books and chapters, author content and publication and institutional contacts. In the biological and life sciences, PubMed Central^{vii}, run by the US National Institutes of Health (NIH), is a free digital archive of the journal literature. Begun in 2000, the archive holds approximately 650,000 items including digitised versions of articles dating back to the 1800s as well as new material added daily.

Institutional repositories, by contrast, have not enjoyed this kind of uptake. OpenDOAR^{viii} is a website listing providing information on over 1,000 academic research repositories. A cursory glance shows that in Australia, institutional repositories contain between a handful and several thousand items, with the larger numbers often representing collections of images, or metadata items without the full access version of a paper attached. This low participation rate in institutional repositories is reflected worldwide. Even at Cornell University, the home of arXiv, academic deposits into the institutional DSpace repository have been low, with faculty indicating that those using a subject archive found it fulfilled their needs, making the institutional repository redundant (Davis & Connolly, 2007).

Various, institutional repositories have been mooted as: a simple way of achieving open access without changing the scholarly communication system or threatening publisher's livelihoods (Harnad, 2003), a method of streamlining university administration systems (O'Brien, 2006), a way to assist with academic workflows (Foster & Gibbons, 2005), or a

¹ A more comprehensive list of subject-based repositories can be found at: <http://www.icbl.hw.ac.uk/perx/analysis.htm> under 3.3 An Analysis of Gap Areas, accessed 1 April 2008

tool with which to fundamentally change the whole scholarly communication system (L. Brown, Griffiths, & Rascoff, 2007; Crow, 2002). There is no doubt that institutional repositories are potentially a very useful tool for many aspects of an institution's administration, from offering a method for collating all the output from an institution, to reporting to funding bodies. Some of the roles of a repository, such as disseminating scholarly communication, overlap with the traditional roles of a university library. Others, such as preserving material and data that is sometimes not usually seen (for example laboratory notes), or negative results that might later be used for analysis, are new to the library's role within the institution. However, while these roles assist administrations and institutions, they do not necessarily assist researchers and there appears to be greater voluntary uptake amongst the research community of subject-based repositories over institutional ones.

The focus of this research

This research investigates how the communication practices between scientists affect the uptake of open access dissemination options in Australia. Currently the coverage of open access journals is not complete across all academic disciplines. The exact number is a 'moving target' but was around 3-5% in 2006 and remains below 10% of the journal market (Willinsky, 2006). Even an academic committed to open access is dependent on either an open access journal being available in their field, or those journals that they publish in adopting a hybrid system that moves closer and closer to a fully open system, or on a publisher making the decision to launch a new open access journal (as PLoS has done).

Repositories on the other hand have enjoyed an explosion of interest and investment in Australia and currently offer a faster (and free to the academic) route to open access. Repositories have been launched in almost every university in Australia, so for Australian researchers to make work openly accessible through a repository is not restricted by availability. For that reason this research focuses mainly on the use of and rollout of repositories.

Publishers and open access

Inevitably there will be some groups in the scholarly publishing system that will be more profoundly affected by substantial change than others. The main arguments against open access come from two actor groups in the debate, commercial publishers and not for profit publishers such as scholarly associations. Commercial publishers argue that open access threatens the viability of publishers, which could lead to a collapse of the scholarly

publishing system, because without the possibility of new journals becoming financially viable it is unlikely publishers will start new titles, and if publishers offer their material as open access or allow material to be placed in repositories, libraries will cancel subscriptions.

Publishers are central to the communication practices of scientists, as all formal communication via scholarly articles must currently be submitted to and published by publishers. Therefore, the approach publishers are taking to the open access issue will to an extent shape some scholar's opinions about the issue. This chapter will now look at how publishers are meeting the challenges that open access has created.

Commercial publishers

Publishers are supporting open access in many ways. Some publishers, under their copyright agreements, allow authors to place their version of a published (or submitted) paper onto a website or into a repository. At first glance, the list of publisher's self-depositing policies^{ix} seems highly supportive of authors self-depositing their work in archives. The 'current journal tally' in September 2008 lists 95% of publishers as 'green' – a description used for publishers who allow self-depositing. There are, however, several caveats to this figure. To begin with there is a distinction between 'full-green' and 'pale-green' with the former representing the 62% of publishers who allow post-print self-depositing and the latter referring to the 29% who allow only the depositing of pre-prints. But this website has (as at November 2008) processed the policies of 457 publishers, which is not the total number of publishers. It is difficult to establish exactly how many publishers there are in the world, but an estimate can be made. In November 2008, the Association of Learned and Professional Society Publishers^x had just under 300 members in November 2008, the Association of American Publishers^{xi} had 260 members and the International Association of Scientific, Technical and Medical Publishers^{xii} had 'about' 100 members. Many publishers will be members of all three associations, but one analysis states the main English-language trade and professional associations for journal publishers collectively include 657 publishers producing around 11,550 journals (Ware, 2006). It is clear that the total number of publishers worldwide is greater than 457.

In some cases publisher's archiving policies have the veneer of opening access to research while actually protecting their commercial interests. An example of one of these policies is Nature Publishing Group, which in January 2005, announced a change to their self-archiving policy: "Authors are encouraged to submit the author's version of the accepted, peer reviewed manuscript to their relevant funding body's archive for release *six months after* publication." (author's emphasis) (Nature Publishing Group, 2008). This embargo

period means this is actually delayed-access rather than full open access. The announcement accompanying *Nature's* statement did not mention open access:

This policy has been developed to extend the reach of scientific communications, and to meet the needs of authors and the evolving policies of funding agencies that may wish to archive the research they fund. It is also designed to protect the integrity and authenticity of the scientific record, with the published version clearly identified as the definitive version of the article.²

An early argument against open access made by the publishers was that there is some sort of inherent danger in there being free and open access to scientific and medical information by the general public. Dr John Jarvis, director of Wiley Europe argued in 2004 that if material were more easily available then the general public might use the information badly:

I will say again; let us be careful because this rather enticing statement that everybody should be able to see everything could lead to chaos. Speak to people in the medical profession, and they will say the last thing they want are people who may have illnesses reading this information, marching into surgeries and asking things. We need to be careful with this very, very high-level information (UK House of Commons Science and Technology Committee, 2004a).

This statement was later dismissed as a “silly one, probably expressed in the heat of the debate and not meant seriously” (Velterop, 2008, p. 118).

The publishers' arguments have since become more sophisticated, with the American Association of Publishers (AAP) hiring a public relations consultant on the issue (Giles, 2007), which led to the launch of the Partnership for Research Integrity in Science and Medicine (PRISM)^{xiii} in 2007. This appears to be an attempt by publishers to ‘market’ themselves as the only medium that can be trusted with a gate-keeping role (Firestone, 2007).

This partnership stated that it intended to alert policy makers and citizens to “the very real threat to peer review that ill-considered government interference represents, and to explore the ways in which we can safeguard peer review as a critical component of scientific integrity” (Firestone, 2007). Many commentators have come out in reaction to this initiative, claiming that open access does not threaten peer review and that PRISM is a

² Information from American Scientist Open Access Forum available at: <http://listserver.sigmaxi.org/sc/wa.exe?A2=ind05&L=american-scientist-open-access-forum&F=l&S=&P=3001> accessed 29 September 2008

thinly disguised lobbying group for the publishing industry (Suber, 2007c). Despite PRISM not publicly declaring who their supporters are, it appears the main organisations behind it are the AAP, Wiley, Elsevier and the American Chemical Society (Giles, 2007).

PRISM has also had negative feedback from the people and organisations it purports to represent, with some groups and commentators specifically asking for clarification from their own publishing groups (Rust, 2007). At least one member of the council of AAP resigned over the issue (Howard, 2007). The Rockefeller University Press wrote to the AAP to ask that they put a disclaimer on the site stating that they were not in any way associated with PRISM³. In the context of the open access debate, the launch of PRISM indicates that the publishing industry is concerned enough about the moves by governments and funding bodies towards mandating open access to take action.

Another argument that is put forward by publishers against the widespread use of repositories for open access content is that as more material becomes open access, libraries will start to cancel subscriptions. There is little evidence to support this argument. Looking at the arXiv example, this highly successful and almost universally used (in the relevant disciplines) repository has been shown to have had no effect on the subscription rates of the journals publishing the final versions of the paper appearing in the repository (Beckett & Inger, 2006; Swan & Brown, 2005). There is evidence to show that readers use the preprint version until it appears in the journal, after which time the journal is where readers turn for the article. This represents “coexistence rather than competition” (Pinfield, 2007). Two studies that have asked the question of subscription cancellations (Beckett & Inger, 2006; Mark Ware Consulting, 2006), found pricing and usage were far more important factors in cancellation than the availability of material in an open access format. Despite the finding that the threat to subscriptions is high charges rather than the author’s ability to deposit articles in repositories, these two studies are being used by the publishing industry to argue that institutional repositories spell a very real threat to publisher’s livelihoods (Morris, 2007b).

Non-profit publishers

Scholarly associations or learned societies are a group of actors who are key players in the scholarly communication system, so will be affected by any change. While exact numbers are difficult to ascertain, approximately half of the journals listed in Ulrich’s appear to be associated with non-profit organisations. Scholarly associations generally publish their own journals, but in some cases association journals are published by the commercial

³ One group calling themselves the “Partnership for Integrity in Scientific Dis-semination” have developed a parody of the PRISM website: <http://pisdcoalition.org> accessed 5 October 2008

publishing sector. The five largest publishers produce just under 14% of the journals owned or sponsored by non-profit organisations (Morris, 2007a, p. 302).

Scholarly associations have a revenue stream from membership fees and (often subsidised) subscriptions to their journals. Associations use any surpluses from membership or subscriptions for several purposes, including keeping conference fees low, providing bursaries for attendance at the organisation's own and other meetings, offering research grants, public education, and keeping membership dues low and generally supporting the running costs of the organisations (Baldwin, 2004). However, it seems that the price charged to members for their society subscriptions is, in general, not covering the costs of providing the print journal, which means there is heavy reliance on institutional subscription revenue to support the journals, while the number of institutional subscriptions is falling (Waltham, 2006).

The open access debate has been particularly relevant to scholarly associations as they exist to advance the professional well-being of their members, including providing access to a forum for advancing knowledge. Given this, it is difficult to argue that keeping the status quo with subscriptions as a source of income is addressing these goals (Willinsky, 2003). This debate has opened questions such as: should scholarly societies primarily be fund-raising organisations for other activities in their disciplines, using their publications to bring in the necessary money, or should they be promoters of efficient scholarly communication and use their publications more directly to that end - for instance, by embracing open access? (Velterop, 2003).

There is no suggestion in these debates that societies abandon their publisher role, as they are arguably best placed to organise Certification, Rewards and the sixth function of Navigation (Armbruster, 2007). However researchers in one study have expressed that a move to open access means scholarly associations may not remain financially viable: "there was deep concern expressed over the financial sustainability of a switch to [the open access] model across the board" (Pitts & Stanley, 2007, p. 26). This is despite the open access model being aligned with societies' missions and providing increased visibility to their journals, and the authors and research they publish. At least one commentator has argued that moving to open access would actually be beneficial to small publishers (many of which are society publishers) because author charges represent a new source of income, and "small publishers will be able to compete for this source of revenue equally with large publishers as it is the impact of the journal and quality of publishing experience for the author that is important, not the ability to bundle large packages of journals" (Prosser, 2004, p. 21).

Recent research by Waltham (2008) has shown that having cheap access to the society journal is not the primary reason for people joining societies:

Numerous surveys show that the primary reason for being a member of a society is for the opportunities that membership brings for conferences, networking and collaboration. The journal is always further down the list, sometimes in second place but never in first. Jobs and grants also usually feature quite highly, but the main reason for joining is the chance to get together with like-minded people (pp. 9-10).

There is an argument that societies would continue without the financial income derived from subscriptions because: "If learned societies are valued by their communities, which we believe to be the case, members are likely to remain loyal irrespective of the publishing model employed by their society" (UK House of Commons Science and Technology Committee, 2004b). This argument has been supported in at least one study where:

all the researchers agreed that non-profit societies serve many necessary functions, in addition to journal publishing, for the fields they serve. Their members would not allow them to falter, and would if necessary support their journal through increased membership dues, if converting to an OA model were to have the effect of reducing society revenues" (Pitts & Stanley, 2007, p. 247).

Both publishers and scholarly associations potentially have challenges they will need to address to be able to continue within an open access system. It is worth noting their responses to the situation, because this could have bearing on how individual researchers perceive open access. As stated earlier, it is the individual scholar who holds the key to widespread uptake of open access, so this chapter will now discuss in some depth how researchers have responded to open access dissemination options to date.

Researchers and open access

Central to the question "How are the communication practices between researchers affecting the uptake of open access scholarly dissemination in Australia?", is the individual scholar. In Chapter 2 it was shown that open access offers a more effective way to disseminate knowledge than the traditional scholarly publishing system. It was also demonstrated that open access is a relatively simple publishing option. This leads to the question, if the scholarly communication system is so problematic and open access offers a better alternative, why has open access not enjoyed widespread uptake?

Scholars engage with the scholarly system on a daily basis, as authors, as readers and as reviewers. They have a vested interest in the system because it underpins their careers.

Any change to the scholarly system will only occur if individual scholars are convinced the change will increase the usefulness, effectiveness and usability of the system. Information on individual views has generally been obtained through surveys in Europe and the United States, with several taking a broad sweep of the international scholarly community.

Awareness of and attitudes towards open access

It is interesting to note that despite several years of active debate and conferences, including legislation changes in the US and a parliamentary enquiry in the UK, “levels of awareness of the kinds of issues that are the focus of publishing seminars and library conferences are really surprisingly low among the research community” (Rowlands et al., 2004a). It is clear there remains a large proportion of the academic community who is unaware of the concept of open access, let alone the initiatives they might be able to use. One international study of 3787 authors showed that 82% of respondents knew ‘nothing at all’ or ‘a little’ about open access. The study points out that this is surprising given that the respondents were a self-selecting group who chose to complete a survey about new developments in journal publishing (Rowlands et al., 2004a).

While *awareness* of open access is limited, there is some indication in the literature that *attitudes* in academic populations reflect open access principles. One study found researchers, “tended to view OA from a ‘philosophical’ perspective – as a fundamental requirement of largely publicly funded research, facilitated by the ease of electronic dissemination” (Pitts & Stanley, 2007, p. 246). Other studies have demonstrated that researchers have a theoretical willingness to publish in open access outlets. In a study of the faculties of Economics and Law of the University of Study of Brescia, Italy, over 66% of respondents answered they were prepared to personally deposit their own scientific or educational material in an institutional repository, once the conditions they requested had been fulfilled (Pelizzari, 2003). These conditions were (in order of numbers of requests); the possibility of continuing to publish in traditional channels, guaranteeing the integrity of their work, and indexation of the work to protect against irretrievability.

A study of 75 researchers has indicated that over two thirds of the respondents do (or would if a repository was available) deposit in institutional repositories (J. Allen, 2005). Another much larger study of 1296 researchers found that 81% of respondents would ‘willingly comply’ if their employer or research funder required them to deposit copies of their articles in an open archive (Swan & Brown, 2005, p. 63). However, these positive attitudes and expressions of intent are not translating into action, and it is important to understand the reasons why.

Barriers to publishing in open access outlets

There is an argument that the demand for open access dissemination is limited. An open access debate that was run through *Nature* in 2003 and 2004 did not end up greatly in favour of the idea (D. Butler, 2004). Another analysis in the same year concluded: "Clearly there is some sort of groundswell, but it certainly was not overwhelming, and early indications from proceedings of National Academy of Sciences in America and such like have not really supported the contention that it is huge" (UK House of Commons Science and Technology Committee, 2004a). This analysis found the groundswell appears to be in the molecular biology end of the science spectrum. This may be because molecular biology had an unprecedented experience with open access when the Human Genome Project^{xiv} was abandoned by traditional publishers and was created as an open access program, which transformed biomedical science.

Given the evidence of increased visibility for open access articles, it might be expected that researchers would be clamouring to be involved, yet as discussed in Chapter 1, the uptake has been low, with approximately 15% of articles being deposited into a digital repository (Poynder, 2005; Sale, 2005b), and only 10% of journals as open access (Lund University Libraries, 2008). However a more recent estimate indicates the numbers are even lower, with 8.1% of articles openly accessible in journals and 11.3% of articles published which are available as copies deposited in e-print repositories or homepage. Combining these two figures gives an estimate of 19.4 % of the total yearly output which can be accessed freely (Bjork et al., 2008). This estimate is an addition of 4.6% of journal articles that were open access for the year 2006 and a further 3.5% which were delayed-access and available after an embargo period of one year, totalling 8.1%.

There is concern that publishing in open access journals does not count for promotions. In a study by Swan and Brown (2004a), "69% [of traditional authors] said they perceived OA journals in their field to have low impact and the same proportion said they perceive the OA journals in their field to have low prestige" (p. 220). This is not, in fact the case. A study of the nearly 200 open access journals that Thomson Reuters covers, found that the impact factors were similar to traditional journals, ranking close to the 50th percentile within their fields (Testa & McVeigh, 2004). In addition, 55% of traditional authors said they thought publishing in an open access journal would adversely affect their chance of winning grants, 74% thought it would limit the potential impact of their work, and only 42% thought it would adversely affect their chance of appointment and promotion.

Encouraging researchers to deposit their work in repositories has proved to be a difficult challenge. It appears there are three main barriers to filling repositories. The first, and largest, is a lack of awareness or understanding either of open access itself, or of how to

participate. The second is a marketing issue. In asking faculty to add to their workload, there appears to have been little consideration to the 'what's in it for me?' question. The third challenge is the difficulties faced by those researchers who are aware and willing to deposit, but do not (or perceive themselves not to) have the necessary skills or expertise to self-deposit.

A likely reason for the low level of deposits in digital repositories is "ignorance or inertia" (Swan & Brown, 2004b, p. 69). The concerns that authors have about depositing material in an institutional repository include technical issues, the submission process, concerns about having preprints in the public domain and concerns about copyright (Pinfield, 2001). In one study, where about a fifth of respondents had deposited scholarly material in an institutional repository, the most published format was theses, followed by conference papers, accepted papers, pre-prints and datasets (Rowlands et al., 2004a). The problem might simply be one of a lack of awareness. The level of awareness of e-print archives was much lower than the respondent's familiarity with open access journals in one study (Swan & Brown, 2004a).

Another reason for the low uptake of repositories could be the need to make repository content more accessible and usable. There are many search engines apart from Google, which is very widely used (Swan et al., 2005), including Scopus, OAlster and BASE, but we are a long way from the holy grail of the 'One Great Scholarly Search Engine' (Willinsky, 2006). One method of encouraging repository use is to integrate them into current workflows, as is occurring in some disciplines. Today, some journals require pre-print submission to be made via a repository and the peer review process is directed through the repository, where reviewers are sent the url of the repository item rather than a file (Advances in Theoretical and Mathematical Physics, 2007).

Several case studies have offered different methods for obtaining material for the repositories, from trawling researcher websites with material and asking permission to transfer these to the repository (Andrew, 2003), to finding out which journals allowed the self-deposit of articles, and tracking which researchers at the institution have published in those journals (Mackie, 2004). Two other approaches have attempted to make the repository more in line with the researchers' natural communication requirements, such as creating personal web profiles for individuals (Foster & Gibbons, 2005) and developing communities for appropriate groups with their own workflow (Chan, 2004). The University of Melbourne has chosen a holistic approach, attempting to tie the UMER repository in with the university's administrative, financial and reporting systems. The aim is that users enter their details once and the information is then available in all

relevant parts of the system. This aims to link funding, research, data, publications, access, citation, impact and assessment (O'Brien, 2006).

An obvious barrier to self-deposit in repositories is the difficulties in the process of placing an article in a repository, including entering the metadata about the article. Metadata describes the article and allows it to be harvested, citation linked, and searched seamlessly as if all papers were in a global archive. Information includes author, title, publication and date. A study analysing the number of 'keystrokes' this process takes found that, "a researcher writing one co-authored paper per month ... spends about 39 minutes per year in metadata entry tasks related to self archiving" (L. Carr & Harnad, 2005, p. 6). This result was based on the conclusion that, on average, the time taken to upload a paper and enter the metadata was 10 mins 40 secs, with an average 1500 keystrokes to self-deposit. In addition, this deposit time shrinks as a user deposits more papers. The implication of this result is that self-depositing articles is a simple matter, and this study is often used by open access advocates in their arguments, however it does not reflect the true situation.

It is reasonable to argue that researchers should not have to take responsibility for depositing material: "Faculty are typically best at creating new knowledge, not maintaining the record of this process of creation ... Most individual faculty lack the time, resources, or expertise to ensure preservation of their own scholarly work even in the short term" (Lynch, 2003). Having an administrative person undertake the depositing process was identified as a positive for a group in one study which showed the respondents objected to the idea of self-depositing, not on moral or conceptual grounds, but on purely practical ones:

A substantial part of the 16% who declared their unwillingness to self-archive do so, not so much as a refusal of the initiative but rather as a request that others carry out the activity of archiving the material produced by the authors (generally, departmental or faculty technical/administrative personnel are mentioned) (Pelizzari, 2003, section 6.5).

Another barrier to adoption generally experienced across most disciplines is a lack of time or technical expertise on behalf of the academic. "All of the options for self-submission assume a basic level of IT literacy ... In any institution there is an enormous range of IT literacy both between and within departments" (Pinfield, 2001). There are difficulties with using proprietary software for items being deposited into a repository for what is intended to be the longer-term (Barnes, 2006b). Issues such as Microsoft Office Word 2007 not being backwardly compatible to previous versions of Word illustrate the difficulties of using this software in a long-term storage capacity. One way of addressing this is to ask authors to convert their documents to a pdf before depositing them. An open-

access software program is currently being developed in Australia to automate these conversions (Barnes, 2006a) but until this is operational and deployed, the alternative is to provide a staff member to assist with the conversion and depositing process.

As mentioned earlier, the issue of the logistics of copyright is one that must be addressed for open access to be successful. As discussed above, many publishers do allow archiving of pre- and/or post- prints and there is a website researchers and administrators can use to determine publisher copyright policies^{xv}, however most researchers appear to be unaware of this. There is also the complicating factor of a changing publishing market, where larger publishers are often buying smaller and independent titles. The self-depositing status of the author of a paper that was published under the imprimatur of a publisher that is now owned by a new company remains somewhat unclear. Self-depositing, while in principle a simple task, can be more complex than it at first appears.

In the case of the University of Toronto, which launched T-Space in early 2003, the decision was made to employ a person 12 hours a week to digitise print documents and converting files into Adobe's pdf format, checking copyright and sending out permission requests to publishers: "The library's decision to perform archiving is intended to maximize the workload of the faculty, to fill the repository quickly, and to learn about the range of issues that may arise as a result of diverse types of submission" (Chan, 2004, p. 288). The T-Space repository contains over 16,000 items in November 2008, and a quick calculation based on the 'browse by date' facility^{xvi} indicates that there was a surge in deposits in 2004 and 2005 with over 2200 items deposited in both years.

Deposit mandates

Several commentators in the open access debate argue strongly for the need to mandate self-depositing at a national or institutional level, rather than relying on individuals to make the decision to do so (Harnad et al., 2004a; Law, 2006; Sale, 2007b). In 2005 several significant funding bodies released policies requiring or requesting grant recipients to deposit any papers resulting from their grants in a publicly accessible digital repository. In October of that year, two mandatory policies came into effect in the UK. The Wellcome Trust, which grants funding that produces approximately 3500 papers a year, stated that "copies of the final manuscripts of all authors' research papers, supported in whole or in part by Wellcome Trust funding, must be deposited in [PubMed Central](#) as soon as possible, and no later than six months after publication" (Wellcome Trust, 2005). This statement is in line with its position statement in support of Open Access Publishing (Wellcome Trust, 2004b). The Research Councils of the UK (RCUK) released a policy effective the same date

stating that all grant recipients should deposit a copy of any resulting work in an appropriate e-print repository (Research Councils UK, 2005b).

The Scientific Council of the European Research Council (ERC) distributes about 15% of the European Union research budget. In December 2007, the ERC (2007) issued a mandate to make both data and articles funded by the ERC open access. This is the first European Union wide mandate. From April 2008, the National Institutes of Health (NIH) in the US has required investigators to deposit their articles stemming from NIH funding in the NIH online archive (Association of Research Libraries, 2008). This field is constantly moving and new initiatives are being announced on a regular basis, such as Harvard University mandating that all research be available in open access format (Darnton, 2008).

Certainly non-mandated requests for repository use do not result in uptake of repositories on a large scale, as demonstrated in the US where a 2005 'request' that researchers funded by the NIH deposit copies of their work into PubMed Central (National Institutes of Health, 2005) was met with limited success, by 2007 only 10,000 of the as many as 65,000 articles derived from NIH-funded research, were available at PubMed Central. Surprisingly, authors sent in only 4% of articles compared to 10-12% submitted by publications (Biello, 2007). The NIH has since upgraded to a mandate requiring deposit within 12 months (Association of Research Libraries, 2008).

As at September 2008 there were five institutional mandates in place in Australia, QUT, Charles Sturt University, Macquarie University, James Cook University and Central Queensland University. In addition, five other Australian universities have indicated they are intending to implement a mandatory deposit policy (Kennan & Kingsley, 2009). In addition, two major funding sources in Australia have recently requested the placement of scholarly output into repositories.

Australian research grants are generally funded through two bodies which "encourage researchers to consider the benefits of depositing their data and any publications arising from a research project in an appropriate subject and/or institutional repository wherever such a repository is available to the researcher(s)" (Australian Research Council, 2007a, p. 13; National Health and Medical Research Council, 2007). This development has been welcomed by commentators (Sale, 2007a). The ARC policy goes further, saying that if the researcher chooses not to do so they must include their reasons why not in their Final Report. To some observers, this requirement for a justification of non-compliance has meant, "this effectively converts the request into a mandate" (Suber, 2007a). While this appears to be supporting open access, in reality, the soft wording used (such as 'encourage' and 'consider') means there is currently little imperative for Australian

researchers to voluntarily place their work in repositories. However there are indications the Australian Government will strengthen these requirements and there has been a commitment on behalf of the ARC to check compliance of the funding rules (Cooke, 2008).

There is a distinction between institutional repositories and subject-based repositories that is relevant to the discussion here. In the case of institutional repositories, the policies on the selection and retention of material, as well as the general scope and organisation of the repository, is determined by the institution. This stands in contrast to the discipline- or subject-based repository where depositing policies are determined by the research communities. These often develop in an 'organic' manner in response to a specific need in a discipline (Chan, 2004). Attitudinal research of law and economics researchers has indicated a preference for subject-based repositories over an inter-discipline based archive (Pelizzari, 2003). In 2006 there were many more subject based repositories than institutional ones: "of the estimated 2,200 such digital repositories, only 11% are based around a specific institution or institutional department and approximately 35% are subject based" (Regazzi & Caliguiri, 2006, p. 188). Within the subset of subject based repositories, PubMed Central and arXiv accounted for 84% of all records in that analysis.

A study undertaken at Edinburgh University analysed the research material already held on departmental and personal web pages prior to establishing an institutional repository, and found a direct correlation between "the willingness to self archive and the existence of subject-based repositories. Most of the academic units that have a high percentage of self-archiving scholars already have well-established subject repositories set up in that area" (Andrew, 2003, p. 12).

In summary, despite open access dissemination options becoming increasingly more available, and the existence of various mandates to deposit in repositories, the level of engagement with open access by researchers continues to be low.

Previous research into researcher engagement

There have been several studies looking at the level of uptake of open access. Some of these studies have tried to establish the level of self-depositing through electronic searches of the literature to determine how much has been made available as open access. For example Antleman (2006) undertook an online search of self-deposited articles that had been published in six social science disciplines over an 18 month period and found that authors are self-archiving according to the norms of their respective disciplines rather than following self-archiving policies of publishers. Bergstrom (2007) searched for open access versions of articles published in a set of economics journals and compared these results to those from a similar exercise in political science, finding that about 90% of

articles in the most-cited economics journals and about 50% of articles in less-cited journals were available, compared to about 30% of the political science articles which were freely available. While useful snapshots of the uptake of repositories, these studies do not answer the question of *why* researchers are making these publishing decisions.

Bjork (2004) in a review article of the topic, concluded that general awareness of the advantages of open access was a prerequisite for scientists choosing to use open access channels, and that branding and creating a critical mass of users would help encourage repository use. This article is an example of much of the current research and literature which focuses on the perceived benefits of institutional repositories.

Other research has been in the form of large quantitative surveys. For example a German survey of over 1000 respondents which was undertaken to determine the acceptance and use of open access publications in the German scientific community asked about publishing habits of researchers and their means of accessing information. The study found few people engage in open access, and many have reservations about both open access itself and about spending their research budgets on open access journal publication, but many expressed an enthusiasm for the concept of open access (Over, Maiworm, & Schelewsky, 2005). There have been several other quantitative surveys on this topic, findings from which have informed the discussion above (J. Allen, 2005; Pelizzari, 2003; Pinfield, 2001; Rowlands et al., 2004b, 2006 #775; Swan & Brown, 2003, 2004 #94; 2004b, 2005).

There has thus been considerable research into the 'how many' and the 'who' questions of open access, and particularly repository uptake. Many of these studies have been conducted as internet surveys, which, like mail and telephone surveys, have the limitation of having answers restricted by the choices on the questionnaire, or leading the subjects to make choices based on the structure of the question (Detlefsen, 1998). To date, there has been little work that attempts to answer the question of 'why' scholars are not engaging with open access, and there is a question whether a survey would be appropriate to answer this question. One exception is research by Houghton et al. (2003), which "focused on why researchers do what they do, rather than simply on what they do, because it is only by understanding the evolving needs of leading researchers that we can effectively resource research activities in the future" (p. ix). That work was an examination of evolving work practices and their implication for scholarly communication and outlined a coherent agenda for the evolutionary development of a scholarly communication infrastructure.

Brown (1985) states that structured questionnaires are a useful research procedure for establishing what people have experienced. In this type of research, the investigator attempts to explain the relationships found in the questionnaire responses. He goes on to argue however, that “enabling the respondent to provide his or her own account of why events occurred may often bring to light aspects that may remain invisible when looked at in a traditional way” (p.220).

There have, nevertheless, been few studies which have taken a qualitative approach to the issue of research engagement with open access. One example is a study which consisted of semi-structured telephone interviews with 28 randomly selected international authors who submitted to the BMJ in 2003. It investigated attitudes towards open access publishing and author charges (Schroter, Tite, & Smith, 2005). This study found that while almost all participants supported the concept of open access, few had submitted to an open access journal and this was because of a combination of a lack of awareness of which journals publish as open access, concern about journal quality and a dislike of the concept of charging authors for publication.

The best estimates of Australian participation in open access come from large international studies which include a small number of Australian researchers (Primary Research Group Inc, 2007; Rowlands et al., 2004a; van Westrienen & Lynch, 2005). These numbers alone do not give any further indication as to why these researchers are generally not embracing open access.

There are some indications that looking at different disciplines would be a useful approach to understanding the problem. One example is an indepth paper by Kling (2000) which refers to the “social practices that support trustworthy communications in diverse scientific disciplines” (p.1315). Given that the research question is concerned with communication practices, the differences in publication behaviour between disciplines is highly relevant. This, however, has not been a focus of much of the literature in this field. “Unfortunately much of the literature about scholarly e-publishing homogenizes the character of publishing” (Kling & McKim, 1999, p. 896). Kling notes that there is little understanding between disciplines of differences in publishing. For example, while humanities disciplines value books as a publication forum, the lab sciences typically devalue book and book chapter publication. This is an important issue, and should be addressed.

It is clear therefore, that the key to the success or otherwise of open access in the future is the individual scholar. In particular, a focus on their disciplinary practices has the potential to contribute to understanding the problem. Differences between disciplines

have been recorded in few instances, and these are elaborated further in Chapter 4. This thesis therefore addresses these issues and informs the open access debate in an area that has, to date, been somewhat neglected.

Summary

Scholarly communication, while a billion dollar business, is ostensibly a tool for scholars to share their research findings, and for research to build upon itself. This tool has, however, become less effective since the Second World War and the subsequent explosion in scientific publishing. The literature indicates that open access offers a solution to the broad criticisms of the scholarly communication system listed in Chapter 1 and to each of the issues identified in Chapter 2. Despite this, the open access dissemination options currently available to scholars, open access journals and repositories, are not being embraced by the academic community.

A broader view of the communication practices of scholars is needed to understand the motivations behind the publication choices they are making. This thesis picks up the argument of Houghton et al. (2003) that:

It will be essential to take an holistic approach to 're-engineering' the system, which treats the creation, production and distribution of scientific and scholarly information, the management of information rights and access, systems of evaluation and the underlying infrastructure as parts of a single research information infrastructure and scholarly communication system (p. 127).

While there have been many studies to date surveying researcher attitudes, and there has been strong aggregate data compiled describing publication practices, there has been very little attention paid to the holistic view of the researcher and how he or she communicates with all members of his or her working community. Until this is established it will not be possible to truly understand a researcher's motivations, and therefore not possible to address the underlying reasons for the to-date low uptake of open access dissemination options. An important and neglected area of exploration is that of disciplinary differences. There is also a lack of information about Australian researchers more generally.

This research therefore addresses this problem by asking: "How are the communication practices between researchers affecting the uptake of open access scholarly dissemination in Australia?" The work is focused on the individual scholar as the key to change. Scholars are the catalyst for, and the providers and users of scholarly communication and any change to the scholarly communication system, such as a move to open access, will need to

be embraced by the scholarly community. The following chapter will describe in detail the research design used in this research.

- ⁱ Website: <http://pkp.sfu.ca/?q=ojs> accessed 5 October 2008
- ⁱⁱ Website: <http://www.openarchives.org/> accessed 22 September 2008
- ⁱⁱⁱ Website: <http://roar.eprints.org/> accessed 22 September 2008
- ^{iv} Website: <http://univ.cc/> accessed 22 September 2008
- ^v Website: <http://www.opendoar.org> accessed 5 October 2008
- ^{vi} Website: <http://repec.org/> accessed 5 October 2008
- ^{vii} Website: <http://www.pubmedcentral.nih.gov/> accessed 5 October 2008
- ^{viii} Website: <http://opendoar.org/> accessed 5 October 2008
- ^{ix} Website: <http://www.sherpa.ac.uk/romeo/> accessed 5 October 2008
- ^x Website: http://www.alpsp.org/ngen_public/default.asp?id=380 accessed 21 November 2008
- ^{xi} Website: <http://www.publishers.org/> accessed 21 November 2008
- ^{xii} Website: <http://www.stm-assoc.org/> accessed 21 November 2008
- ^{xiii} Website: <http://www.prismcoalition.org/principles.htm> accessed 5 October 2008
- ^{xiv} Website: http://www.ornl.gov/sci/techresources/Human_Genome/home.shtml accessed 5 October 2008
- ^{xv} Website: <http://www.sherpa.ac.uk/romeo/> accessed 5 October 2008
- ^{xvi} Website: <https://tspace.library.utoronto.ca/browse-date> accessed 21 November 2008