

Those who don't look don't find: Disciplinary considerations in repository advocacy

Danny Kingsley
Centre for the Public Awareness of Science
Australian National University
Acton ACT 0200
danny.kingsley@anu.edu.au

This paper has been accepted with minor editorial changes for *OCLC Systems and Services: International Digital Library Perspective (OSS:IDLP)*, Vol 24, No 24, early 2009. This will be a special issue related to the areas of open access and scholarly communications (including case studies, scholarly communication case studies)

Abstract

Purpose of this paper

By describing some of the often ignored aspects of repository advocacy, such as disciplinary differences and how these might affect the adoption of a particular institutional repository, this paper aims to offer practical guidance to repository managers and those responsible for open access and repository policy.

Design/methodology/approach

The argument uses examples from an empirical study of 43 in-depth interviews of academic staff in three disciplines, Chemistry, Computer Science and Sociology, at two Australian universities. The interviewees discussed their interaction with the literature as an author, a reader and a reviewer.

Findings

Disciplines are markedly different from one another, in terms of their subject matter, the speed of publication, information seeking behaviour and social norms. These all have bearing on the likelihood a given group will adopt deposit into an institutional repository as part of their regular work practice.

Practical implications

It is important to decide the purpose of the institutional repository before embarking on an advocacy program. By mapping empirical findings against both diffusion of innovations theory and writings on disciplinary differences, this paper shows that repository advocacy addressing the university academic population as a single unit is unlikely to be successful. Rather, advocacy and implementation of a repository must consider the information seeking behaviour and social norms of each discipline in question.

What is original/value of paper

The consideration of disciplinary differences in relation to repository advocacy has only begun to be explored in the literature.

Introduction

The widespread uptake of the internet in the scholarly world over the last 15 years offers opportunities to reform the long-standing scholarly communication system. Repositories have been mooted as a way to achieve open access, amongst other possible uses, but to date, particularly in institutional repositories, deposit of material

has been slow. This paper is looking at the challenges facing digital repositories in facilitating open access and how they are changing scholarly communication. In such a discussion it is necessary to explore what repositories are intended for. This depends on not only the type of repository in question, but whether the end-user is an institution, an academic researcher, a practitioner or the general public. Given the work of academic researchers is usually the intended content of these digital repositories, we will take an in-depth look at the work practices of these researchers to determine the barriers to a general embracing of repositories as part of the scholarly communication process.

This paper will examine the introduction of repositories into the academic environment in terms of diffusion of innovations theory before discussing disciplinary differences and how this affects the acceptance or not of repositories by certain academic groups. Specifically, the information-seeking behaviour within disciplines has a direct bearing on the likelihood of a given group to accept repositories as they are currently structured. Throughout, examples will be given from a research project into three disciplines based at two universities in Australia. The conclusion makes recommendations to institutional repository managers for achieving a more enthusiastic uptake of their repository.

The open access argument

The open access movement has been active for over a decade. Broadly advocating that peer-reviewed scholarly material should be freely available on the internet at the time of publication, the movement originally developed from a reaction to the scholarly 'crisis' of the 1990's when journal prices skyrocketed (Harnad, 2003). Exact definitions of what constitutes open access have since been determined (Open Society Institute, 2002; Max Planck Institute, 2003).

There are generally two ways to achieve open access currently: open access publishing and using a digital repository to deposit the author's version of an article at the time of submission or publication. These are referred to as the 'gold' and 'green' roads to open access respectively (Harnad et al., 2004). Open access publishing has historically been in specifically created open access journals, such as *PLoS Biology*, or in journals that have moved from a subscription-based model to an open access model, such as the *Medical Journal of Australia*. Generally open access journals are funded either through a pay-on-acceptance charge (sometimes inaccurately referred to as author-charges), or through scholarly association membership fees. It should be noted, however, that most open access journals will waive the charge for authors who are unable to pay.

In the last two years, the 'hybrid model' has become increasingly popular with publishers, who offer authors the opportunity to have their article freely available at time of publication for a fee. Some of these journals 'anticipate' 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 paper is concerned with the second, 'green' method of achieving open access, making author's versions of articles available online. This can be through an author's own website, although generally repositories are considered to be more 'robust' and searchable due to a requirement that they comply with the Open Access Initiative (OAI) Protocolⁱ which requires interoperable standards for searching of repositories.

Types of repositories

The recent widespread uptake of repositories in institutions (van Westrienen & Lynch, 2005) has largely been due to the availability of open-source software that

offers institutions the ability to build a repository 'out of the box'. The most widely used repository platforms worldwide are ePrintsⁱⁱ and DSpaceⁱⁱⁱ. The release of ePrints by Southampton University in 2001 and DSpace in 2002, jointly developed by MIT and Hewlett Packard, allowed any group or institution to build a digital repository at minimal cost. There are other open-source platforms, and some service providers offer proprietary software platforms, but these will not be discussed here in any depth.

The intended purpose of a repository not only determines the usefulness of a given repository platform as a way of achieving open access, but also affects how easy the repository is to use, and what tools are available to the repository manager to encourage repository use. EPrints and DSpace have been designed with different purposes in mind. The goal of ePrints was to allow for the deposit of author pre- or post- prints to facilitate open access to the material without the reader having to pay a subscription fee. DSpace has a wider remit than ePrints, archiving a range of digital content including images, datasets and other forms of scholarly output (Nixon, 2003).

There is a distinction between institutional repositories and subject-based repositories that is relevant to the discussion here. In the former, the policies on the selection and retention of material, as well as the general scope and organization 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 academics has indicated a preference for subject-based repositories over an inter-discipline based archive. (Pelizzari, 2003).

As a demonstration of this preference, it is instructive to look at the participation levels of three subject-based repositories. The most obvious example is arXiv^{iv}, which was developed in 1991 as an archive for preprints in physics by Paul Ginsparg and hosted at the Los Alamos Laboratory. Now hosted at Cornell University, the scope has expanded to include astronomy, computer science, mathematics and other areas. 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^v is a repository disseminating research in economics, where participants can deposit material through their own institutional repository, or directly to the repository. RePEc's website states that it holds: 222,000 working papers, 316,000 journal articles, 1,500 software components and numerous listings for books and chapters, author content and publication and institutional contacts. In the biological and life sciences, PubMed Central^{vi}, run by the US National Institutes of Health (NIH), is a free digital archive of the journal literature. Begun in 2000, the archive holds digitised versions of articles dating back to the 1800's as well as new material added daily. The archive holds approximately 650,000 items, with most of the recent content added by researchers who have been funded by the NIH.

Institutional repositories, by contrast, have not enjoyed this kind of success. OpenDOAR^{vii} is a website listing and 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 (Ware, 2004a; Pelizzari, 2003; Allen, 2005). 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).

What are repositories for?

Various, 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, a way to assist with academic workflows (Foster & Gibbons, 2005), or a tool with which to fundamentally change the whole scholarly communication system (Crow, 2002; Brown, Griffiths, & Rascoff, 2007). Given that digital libraries will generally be hosting institutional repositories, this paper will now focus on them rather than subject-based repositories.

One of the reasons for the low participation rate in institutional repositories is an issue of purpose. Currently, of the above potential uses of institutional repositories, only the institutional goals of creating a university administration system and a digital library are being achieved. Certainly the large numbers of image and video items in institutional repositories indicate there was a need for this type of facility not previously being met. There are, however, serious long-term problems of sustainability for digital repositories (Bellekom, 2004). This is also an issue for data repositories (Buchhorn & McNamara, 2006), a topic for discussion elsewhere.

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. In some respects, it is not surprising that institutional repositories benefit the institution. Certainly the nomenclature has indicated to the academic community that the repository is designed to support and highlight the achievements of the institution rather than provide any benefit to them (Foster & Gibbons, 2005). But the issues are more complex than a matter of terminology.

Using institutional repositories as a method of achieving open access has, to date, been only partially successful. While it is extremely difficult to quantify not only the number of items freely available in repositories but also the number of articles produced in a given year (Tenopir, 2004), a widely mooted figure is that approximately 15% of published articles are available in open access form in repositories (Sale, 2005).

The broader question of whether repositories are reforming the scholarly communication landscape is well beyond the scope of this article. Suffice to say that while arguments abound that the days of the scholarly journal are limited (the subject of a previous paper (Kingsley, 2007a)) the scholarly communication system is currently deeply embedded in the reward system used in academia (Steele, Butler, & Kingsley, 2006), and until this changes there is unlikely to be a revolution.

The practice of putting author's versions of papers into repositories, despite concerns on behalf of publishers, has so far had little impact on subscription rates. 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).

Let us turn our attention to why repositories are having the successes and failures that they are. In doing so, the argument will now draw on research that has looked at

implementing new ideas into groups of people, described as the diffusion of innovations.

Diffusing repositories into the academic community

In 1962, Everett M Rogers wrote a book, *Diffusion of Innovations*, outlining a new theory of how innovations came to be accepted by groups of people. The 5th edition was published in 2003. Institutional repositories clearly represent a new innovation, defined by Rogers as “an idea, practice, or object that is perceived as new by an individual” (Rogers, 2003) (p.12). The diffusion process is concerned with communication of a new idea to members of a social system, described as “a set of interrelated units...engaged in joint problem solving to accomplish a common goal” (p.23). The implementation of repositories into the academic community fits neatly into these definitions, and this section of the paper will discuss insights from diffusion theory which may help guide those responsible for repository advocacy.

One of the reasons why subject-based repositories have enjoyed relative success could be explained by their focus on a particular discipline. Years of research into diffusion of innovations have demonstrated that diffusions are more successful if managed as a decentralised system, where the participants can make decisions about the diffusion process and create and share information with one another to reach a mutual understanding. Decentralised systems are likely to fit more closely with the user’s need and problems. By contrast, institutional repositories are, by definition ‘centralised’ systems, where the decisions about the innovation itself and the diffusion of the innovation are imposed from an external source – the university administration. In these instances the innovation is diffused as a “uniform package to potential adopters who accept or reject the innovation. The individual adopter of the innovation is thought of as a relatively passive acceptor” (Rogers, 2003, p.395).

To simplify Roger’s argument, an innovation is more likely to be adopted if: the adopter perceives the innovation to be more advantageous than the idea or process it supersedes, if it is consistent with existing values, past experiences and needs of the adopter, if it is perceived to be difficult to understand and use, if it can be experimented with and if the results of the innovation are visible to others. Institutional repositories face difficulties on all these counts to varying extents. Issues of perceived complexity, and demonstrability depends partially on the software platform the repository is built on and how the institution has customised their own repository. For example ePrints offers download statistics for individual papers, and a simple-to-use deposit interface. DSpace has partially attempted to address the disciplinary difference issue by structuring the repository so it reflects ‘communities’ within the university. These can be mapped to the departments of the university.

In a university or other institutional environment, it is fair to say that generally the repository has been developed with the institutional structure in mind (Chan, 2004) Often attempts to encourage repository use have involved university-wide strategies, such as mining personal websites for material academics are already putting online, finding out which journals allow deposit of post prints and approaching authors who have published in them, or determining which OA journals people have published in (Mackie, 2004). But these suggestions, while likely to be effective in the initial goal of partially filling the repository, are heavily reliant on having a centralised person or system in the institution to manage this ingestion. These methods are unlikely to spontaneously encourage widespread use on the behalf of the academic community itself, not least because the academic community is not homogenous, discussed in greater depth below.

There are several commonly encountered problems with the adoption of repositories by academics in universities. One is a matter of language. For example the expression 'post-print' is one used widely within the open access community and within library circles. The general academic community, however is not familiar with this term, so using the expression 'the final corrected post peer review draft version' is far more effective (Callan, 2007).

Another barrier to adoption generally experienced across most disciplines is a simple technological issue, the format of the item being deposited. There are problems 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. A simple way of addressing this is to ask authors to convert their documents to a pdf before depositing them. However while this may seem to be a simple instruction to some people, it can cause difficulties within the general population that may not be as computer literate as assumed and may not have access to the appropriate software. An open-access software program is currently being developed to automate this system (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.

The third issue is one of copyright. While many publishers do allow archiving of pre-and/or post- prints there is a website academics and administrators can use to determine publisher copyright policies^{viii}, most academics are unaware of it. This is reflective of the broader phenomenon that awareness of the copyright status of published work varies in the academic community as does willingness to comply. Checking the copyright status of articles is time consuming and potentially confusing for academics, and is more efficiently dealt with at an administration level (Callan, 2007; Mackie, 2004).

While the above challenges are experienced across the board in academic environments, it would be foolhardy to think of the research community as a homogenous group. The difficulty with developing diffusion policy within an institution is that the 'existing values, past experiences and needs' of academics change according to the discipline. Rather than a single social system, academics consist of a series of small, disparate groups with distinct differences. It is for this reason that a uniform advocacy or 'roll-out' program for a given institutional repository is unlikely to succeed. With this in mind we now turn our attention to disciplinary differences and how they might affect the adoption of repositories.

The disciplinary difference issue

To say that disciplines differ from one another is a truism, however, the extent to which they differ, not only between disciplines but also within them is the subject of this section of the paper. In order to illustrate some of the propositions put forward here, examples will be given from interviews conducted as part of a research project into the barriers to the uptake of open access in Australia. A total of 43 in-depth interviews were conducted at two Australian universities, the Australian National University and the University of New South Wales, from October 2006 to March 2007 with academics in the field of Chemistry, Sociology and Computer Science. The semi-structured interviews discussed the behaviour of the researcher as a reader, a writer and a reviewer of articles as well as canvassing views on open access and attitudes to their institutional repository. After analysis of the transcripts, two interviews were conducted as triangulation, at Queensland University of Technology with the repository manager and the Deputy Vice-Chancellor of Technology,

Information and Learning Support. QUT was chosen because it is the only university in Australia with a mandate to deposit scholarly output into the institutional repository (QUT, 2004). The full methodology of the research project is detailed elsewhere (Kingsley, 2007b).

In choosing the three disciplines for interview, the initial consideration was for the way the disciplines publish their work. Chemistry, representing a hard science, traditionally publishes in peer-reviewed articles in journals. Sociology, while also publishing in this manner, also has a tradition of publishing books or monographs, while Computer Science primarily uses conference proceedings for peer-reviewed communication. Publishing output, however is only one manifestation of the fundamental differences between disciplines, and results from the general 'speed' of the endeavour in question.

Fast moving research with many people working on similar topics is described as urban (using the analogy of urban life) (Becher & Trowler, 2001). High-energy physics and computer science are obvious examples, but the current race for priority in stem cell research identifies this as an urban area. Urban areas of research need a fast form of communication, and the development of repositories like arXiv was merely an electronic extension of an already thriving pre-print culture (Hagstrom, 1970). This hectic pace demands more informal forms of communication. Crane observed as early as 1972 that physicists placed a higher ranking on informal sources of information using sources like conversation and correspondence compared to chemists.

Of the three disciplines interviewed, computer science is the fastest moving. The use of conferences as a method of communicating ideas is the most efficient in this context. Sociology, by contrast fits squarely in the category of rural research, where an individual researcher may be the only person world-wide working on a given topic. Books are an appropriate format for publication in this context. Many of the people interviewed in sociology described delays in journal article publication of two years, in one case an interviewee had been waiting for publication of a book chapter for nine years (although this is not typical it does illustrate how protracted the process can be). These general time frames have been reported elsewhere (Becher & Trowler, 2001). Chemistry falls in the middle with academics in different sub-disciplines reporting a range of publication times.

While generally academics can be described as people who work with ideas, the nature of the particular intellectual tasks on which specific groups are engaged determines to some extent their 'culture'. The divide between disciplines is not limited to the subject being explored. It extends to all aspects of the research endeavour, the language used, the methods of communication and the sources of information, to name a few. If reconsidering the likelihood of adoption of a new technology, such as a repository, the level of engagement a particular group will have towards a technology will be partially determined by their current work practices, and these differ from discipline to discipline. Disciplines themselves are hard to define, but to be admitted to membership of a section of the academic profession "involves not only a sufficient level of technical proficiency in one's intellectual trade but also a proper measure of loyalty to one's collegial group and of adherence to its norms" (Becher & Trowler, 2001, p.47).

Identifying differences between disciplines may not be enough to determine successful ways of implementing repository use, as disciplines themselves encompass a series of sub-specialisms. Many of the computer scientists spoken to made the comment that they were 'unusual' because they 'straddled' another area.

While these areas were all different from one another, the trend of 'straddling' appeared to be almost universal, and certainly within the cohort of computer scientists/engineers interviewed it would be difficult to identify a 'typical' or representative one. This observation has been made elsewhere:

There is no single method of enquiry, no standard verification procedure, no definitive set of concepts that uniquely characterizes each particular discipline. It is in some contexts more meaningful to speak about the identifiable and coherent properties of subsidiary areas within one disciplinary domain or another. (Becher & Trowler, 2001) (p65)

Generally the academic population is unaware of how other disciplines function, "academics seem to be surprisingly hazy in characterising other people's subjects of study, and their stereotypes of both subjects and practitioners are in general neither particularly perceptive nor particularly illuminating" (Becher, 1981, p.110). It can be argued that the university administration is similarly hindered in its understanding of the myriad of work practices and social norms in disciplines. One work practice of relevance in this debate is information-seeking.

Desperately seeking information

The way a given group of researchers search for information has a great bearing on their perceived attitude towards the usefulness of institutional repositories. Generally speaking, researchers undertake two kinds of searching of the literature, broad and specific (Back, 1962). While the term 'keeping up with the literature' might be considered quaint in some disciplines and irrelevant in others, it is still a practice undertaken in defined areas such as chemistry, although techniques have changed with the advent of the internet:

"I used to on Friday morning check all the journals. In the old days we would go to the library"

Several interviewees made similar comments. Now these searches are conducted electronically:

"I get abstracts of journals sent – keeping up with it all is hard. I am on email lists...I look at journals online."

More commonly researchers will be looking at a specific topic, because they are reviewing a paper and wish to ensure that the topic has not been covered elsewhere, or because they are writing a paper on the topic and need to ensure that they have seen or are aware of all other work in the area. It is this latter type of searching that is of most relevance to this paper. By looking at the specific tools different groups of researchers use to find information, clues can be found as to the usefulness or not of a repository to that group. Taking chemistry as the first example, those interviewed indicated that they use a series of tools including SciFinder, Thompson Scientific's Web of Science and Chemical Society Abstracts. There was not a great reliance on Google as a search engine, with a preference for databases.

The chemists, when asked about whether they would place material in a repository, made comments such as:

"I as a user would like something that's searchable not just for an institution but across all institutions"

"My view is it would just get buried, people wouldn't look for it."

Of course, the idea of the repository is that the searcher does not need to go to the institutional web page, they can use a search engine such as Google or OAlster and find the paper, almost without knowing they have found their goal through a

repository. But this mis-perception that items in a repository would not be found by other people reflects the way chemists currently search for information: going to the database where information is housed rather than conducting general searches.

The sociologists in the sample used a wide range of tools to help them with their information search. In keeping with the rural nature of the endeavour, the concept of 'keeping up with the literature' was not adhered to as there is not necessarily a specific literature in a given area of enquiry:

"I am a bit of a generalist in my approach. What it gets down to is largely a matter of accident."

"The ideas are interdisciplinary, the field is so broad I don't worry about covering it"

In computer science, far more so than the other disciplines interviewed, it is common practice to have a personal website with all published papers listed on that site. In many cases there is a version of the paper attached to that listing. This practice reflects the ecological approach advocated by (Gandel, Katz, & Metros, 2004) who suggest personal digital repositories which can then be collated. When asked about the copyright status of those papers, the interviewees indicated that either they thought they had permission, or they were not concerned about potential repercussions from the publishers:

"I haven't asked permission [to put pdfs on my site] but I have had no problems"

"I don't worry about copyright policies"

"All my stuff on the web probably contravenes the lettering of copyright... Publishers aren't bothered about you putting up papers on your website as long as that's all".

Putting the copyright implications of this practice to one side, having information available in this form means that without exception, the computer scientists spoken to used Google as a search tool amongst other methods. The subject of their searches was a person rather than a topic, and the first place to look was an individual's website where the relevant paper (or one that was close enough) could be downloaded.

This last situation is an interesting conundrum for an advocate of an institutional repository. Those researchers who put their papers into personal websites are already practising open access. All the material they use is available freely online via a Google search. Using personal websites might not address some of the sustainability issues that repository developers are trying to resolve, but in a fast-moving discipline, most material is out of date very quickly so this is not necessarily a priority:

"Because I am researching the web – it's changing everyday. If my results are not out in one year ... it will go nowhere"

"Computing moves so fast".

There are evidently in some cases, serious copyright issues with this practice that should probably be addressed for the researchers, but if the institution's focus is on achieving open access, then energy would be better spent, in the case of computer science at least, addressing the copyright problem rather than trying to encourage those academics to alter their behaviour and use the institution's repository.

The mandate argument

When encouraging self-archiving, subject-based repositories have great advantage over institutional ones, “it seems there is a direct correlation between 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). One way of enforcing an increase in use of institutional archives is to mandate deposit into them.

Several open access advocates have called for mandates to encourage repository use (Harnad, 2006), (Sale, 2007). In theory, this is supported by attitudinal studies showing that 80% of academics would willingly place their work into a repository if required to do so (Swan & Brown, 2004). But as the QUT experience shows mandating alone is not the solution. Care must be taken to address both the broader and discipline-specific issues when rolling out the repository (Cochrane & Callan, 2007; Allen, 2005).

Open access as a sales pitch

There is a distinction between an individual’s attitude towards an idea and their behaviour towards it, and open access is an example of this. Those interviewed generally expressed open access sentiments suggesting the results of science should be freely available:

“Research is pretty meaningless if you can’t communicate it. The whole purpose of research rests on disseminating the research”

“I believe work should be published. We are financed by the tax payer, it should be in the public domain”.

“What’s science for if you don’t have things available?”.

However when asked about changing their behaviour such as using a repository, there was less enthusiasm and in some cases antipathy towards the suggestion:

“I can’t see the point of putting thesis on Digital Thesis when I have a copy on my own website”

“I don’t know what benefit it is for me, it sounds like more work to do it”

“I don’t see any harm in depositing in a IR, but don’t see any use in it either”

“I have a concern about plagiarism”

“There are all sorts of copyright restrictions”.

Certainly other studies have shown that in theory, academics support open access (Swan & Brown, 2004) but their practice does not bear this out when looking at what scholarly output is available in an open access format worldwide.

This apparent dichotomy could be for several reasons, not least the method of the introduction of the technology, discussed in this paper. Another compelling reason for resisting changes to their current work practices is scholarly publishing is tied to the reward system in academia, and any change to the practice potentially jeopardises the academic’s standing (Steele, Butler, & Kingsley, 2006; Bjork, 2004; Harley, Earl-Novell, Arter, Lawrence, & King, 2007). In addition, there have been potential clinical concerns expressed about publishing non-peer reviewed articles in chemistry and biomedicine, as well as fear of plagiarism in some humanities areas (Ware, 2004b).

The open access message is not necessarily a good 'selling point' to academics as a reason to put material into a repository. Researchers based in institutions in first world countries already have 'open access' to much of what they require because their institution subscribes to it. Access is not necessarily an issue for them. In the two (well resourced) universities where interviews were conducted, the only access issue expressed was by some of the sociologists who found they needed to buy their own books. Generally however this seemed to be accepted by the interviewees. Having books available in open access form is not what is being discussed here, so for the purposes of this paper, the academics who are being asked to make their work freely available are likely not to be having issues themselves with obtaining the material they use.

Open access encompasses more than simply scholarly communication, which implies communication between scholars, potentially a very private conversation. Academic endeavour is in many ways a social activity (Crane, 1972). Particularly in the sciences, research builds upon itself as researchers report small steps in the movement towards an answer to a large problem that many people are working on. Newton's famous quote 'if I have seen further it is by standing on the shoulders of giants' is a lyrical description of this phenomenon.

In order for this progression to occur, it is essential for researchers to communicate their findings to one another. Traditionally this has been by publishing articles in peer-reviewed journals, but as communication channels have improved, some disciplines have adopted faster, more informal methods of communication. Specifically, the introduction of the internet (which, it can be argued, represents a seismic shift in communication in the order of that of the printing press) has allowed for new types of communication previously unimagined. These 'Web 2.0' techniques, such as blogs, wikis, Skype (to mention a few) are being adopted by many of the computer scientists interviewed.

A few good friends

Generally academic circles are very small, with an immediate group of approximately 5-20 people. A larger group of interested researchers might encompass about 200, but that is the extent of people who would have a direct research interest in an individual's work (Becher & Trowler, 2001). The intimate nature of these groups means researchers are known to one another:

"I follow ...leads given by people I know. I rely on personal networks"

"I know most of the people active in my field, they send me their work.
About 12-20 people".

If we consider the small size of the members not only of one discipline but of the sub-speciality that makes up a particular individual's inner circle, the likelihood is that these people are not working in the same institution, indeed many of the people interviewed discussed their collaborators overseas. Given the requirements in a university environment of covering a broad range of topics for undergraduate teaching, it is not surprising that many academics find their research colleagues outside their own institution (Becher, 1981; Foster & Gibbons, 2005). Academics need to communicate and share thoughts with their small inner circle, and using a tool developed by the institution is unlikely to be the first choice.

Considering the small size of the intended audience of a particular piece of work, it is not surprising that many scholarly papers are never cited. A core of approximately 2,000 journals now accounts for 95% of cited articles (Steele, Butler, & Kingsley,

2006). Even allowing that citation counts are a blunt way to determine how many people read a paper, the academic audience for scholarly papers is not huge. But if we move from scholarly communication and turn to open access the audience becomes considerably broader. There is a large literature demonstrating that having articles open access increases the citations for those papers (Hitchcock, 2006).

Apart from researchers in the third world, there is a wide audience for scholarly output including practitioners, such as teachers, nurses, doctors, medical and scientific lawyers and accountants who work in fields that benefit from research but are usually not in a workplace that subscribes to the relevant journals. Field researchers – for private organizations and for government departments – are also similarly disadvantaged. These are the benefactors of having material available as open access. It follows, then, that institutions with these cohorts might be at an advantage when implementing their repository. Indeed, this was suggested as a possible reason for QUT's relative success in repository uptake (Callan, 2007). It is perhaps surprising then, that the author who consistently heads the Top 50 Authors list in QUT ePrints is researching and publishing in chemistry. With over 61,000 downloads of his papers in the previous year as at (November 2007), a possible explanation for this extraordinary interest is some of the work is in the area of environmental chemistry – another area where there are many field practitioners not tied to institutions.

This current wider audience for scholarly articles does not necessarily translate into quantifiable 'points' for the researcher in the form of citations. The open access argument will only tie back into reform of the scholarly communication situation if it reflects the reward system. If the way 'success' or 'impact' is measured changes (such as a count of downloads of material, for example), then the arguments for making material open access will become considerably more compelling for the academic.

Conclusion

A repository manager, faced with the challenge of encouraging repository use must consider several aspects. While touting the repository as a means to achieve open access may appeal to some academics, the more pressing issues of disciplinary norms and their expected reporting behaviours will take precedence. Addressing these concerns will be the first step in successful repository advocacy. However advocacy alone will not always translate into action by the academic community, and consideration of disciplinary differences when offering reasons and methods for using the repository will ensure a much smoother transition.

Steps such as simplifying the process, offering assistance with the more technical aspects of depositing papers, having a person available on the telephone rather than an email enquiry have all been shown to increase enthusiasm for the repository (Foster & Gibbons, 2005). Adding benefits such as an individual researcher page or tying the process into already existing administration to avoid repeated reporting will encourage take-up of the system because it offers a benefit to the researcher.

When developing a university policy on open access and/or institutional repository use, the existing behaviours of the academic community expected to use it need to be considered. If the purpose of the repository is to achieve open access for the university or institutional output, then those disciplines where open access is already being practiced should be a low priority. Those disciplines unlikely to use repositories to find information will need to be given other reasons why their work should be

made freely available. The concerns of researchers in other disciplines, such as about plagiarism, need to be taken seriously and addressed.

Repositories are unlikely to solve scholarly communication issues in the short term. If open access is a priority, even with a strong open access policy, a mandate and staff dedicated to the process, the initial increase in open access to the institutions output is likely to be slow. If, on the other hand, the purpose of the repository is an administrative one, to assist the institution with reporting for funding, or as a showcase of the university output, then the onus of spending time adding to the repository and maintaining it should fall squarely on the shoulders of the university administration and not the academic community. It is a simple question – who benefits?

Bibliography

- Allen, J. (2005). Interdisciplinary differences in attitudes towards deposit in institutional repositories Retrieved 20 January, 2006, from <http://eprints.rclis.org/archive/00005180/>
- Andrew, T. (2003). Trends in Self-Posting of Research Material Online by Academic Staff. *Ariadne*(37).
- Back, K. W. (1962). The behaviour of scientists: Communication and creativity *Sociological Inquiry*, 32, 82-87.
- Barnes, I. (2006a, 2 February). *Integrating the Repository with Academic Workflow*. Paper presented at the Open Repositories 2006, Sydney University.
- Barnes, I. (2006b, July 2006). Preservation of Word-Processing Documents. Retrieved 30 September, 2006, from http://www.apsr.edu.au/publications/preservation_of_word_processing_documents.html
- Becher, T. (1981). Towards a Definition of Disciplinary Cultures. *Studies in Higher Education*, 6(2), 109-122.
- Becher, T., & Trowler, P. R. (2001). *Academic Tribes and Territories* (Second ed.): The Society for Research into Higher Education & Open University Press.
- Beckett, C., & Inger, S. (2006). *Self-Archiving and Journal Subscriptions: Co-existence or Competition? An International Survey of Librarians' Preferences*. London: Publishing Research Consortium.
- Bellekom, C. (2004). Building preservation functionality in a digital archive: the National Library of the Netherlands. *Learned Publishing*, 17(4), 275-280.
- Bjork, B.-C. (2004). Open access to scientific publications - an analysis of the barriers to change? *Information Research: an international electronic journal*, 9(2).
- Brown, L., Griffiths, R., & Rascoff, M. (2007, 26 July). University Publishing In A Digital Age. *Ithaka Report* Retrieved 28 November, 2007, from <http://www.ithaka.org/strategic-services/Ithaka%20University%20Publishing%20Report.pdf>
- Buchhorn, M., & McNamara, P. (2006, September). Australian eResearch Sustainability Survey. Retrieved 28 November, 2007, from <http://dspace.anu.edu.au/handle/1885/44304>
- Callan, P. (2007). Interview at QUT. In D. Kingsley (Ed.) (pp. Interview). Brisbane.
- Chan, L. (2004). Supporting and Enhancing Scholarship in the Digital Age: The Role of Open-Access Institutional Repositories. *Canadian Journal of Communication*, 29, 277-300.
- Cochrane, T., & Callan, P. (2007). Making a Difference: Implementing the eprints mandate at QUT. *International Digital Library Perspectives*, 23(3), 262-268.

- Crane, D. (1972). *Invisible Colleges: Diffusion of Knowledge in Scientific Communities*. Chicago: The University of Chicago Press.
- Crow, R. (2002). *The Case for Institutional Repositories: A SPARC Position Paper*. Washington: The Scholarly Publishing & Academic Resources Coalition.
- Davis, P. M., & Connolly, M. J. L. (2007). Institutional Repositories: Evaluating the Reasons for Non-use of Cornell University's Installation of DSpace. *D-Lib Magazine*, 13(3/4).
- Foster, N. F., & Gibbons, S. (2005). Understanding Faculty to Improve Content Recruitment for Institutional Repositories. *D-Lib Magazine*, 11(1).
- Gandel, P. B., Katz, R. N., & Metros, S. E. (2004, March/April). "The Weariness of the Flesh": Reflections on the Life of the Mind in an Era of Abundance. *Educause Review*, 40-51.
- Hagstrom, W. O. (1970). Factors Related to the Use of Different Modes of Publishing Research in Four Scientific Fields. In C. E. Nelson & D. K. Pollock (Eds.), *Communication Among Scientists and Engineers* (pp. 85-124). Lexington: Heath Lexington Books.
- Harley, D., Earl-Novell, S., Arter, J., Lawrence, S., & King, c. J. (2007). The influence of academic values on scholarly publication and communication practices. *Journal of Electronic Publishing*, 10(2).
- Harnad, S. (2003). For Whom the Gate Tolls? How and Why to Free the Refereed Research Literature Online Through Author/Institution Self-Archiving, Now. In D. Law & J. Andrews (Eds.), *Digital Libraries: Policy Planning and Practice*: Ashgate Publishing.
- Harnad, S. (2006). FRPAA and paying publishers to self archive. Retrieved 28 November, 2007, from <http://www.library.yale.edu/~llicense/ListArchives/0606/msg00165.html>
- Harnad, S., Brody, T., Vallieres, F., Carr, L., Hitchcock, S., Gingras, Y., et al. (2004). The green and gold roads to Open Access. *Nature Web Focus: Access to the Literature* Retrieved 03 March 2005, from <http://www.nature.com/nature/focus/accessdebate/21.html>
- Hitchcock, S. (2006). The effect of open access and downloads ('hits') on citation impact: a bibliography of studies. *OpCit Project* Retrieved 4 September, 2006, from <http://opcit.eprints.org/oacitation-biblio.html>
- Kingsley, D. (2007a). The journal is dead, long live the journal. *On the Horizon*, 15(4), 211-221.
- Kingsley, D. (2007b). The one that got away? Why changed reporting requirements will work against open access in Australia. *First Monday (Submitted October 2007)*.
- Mackie, M. (2004). Filling Institutional Repositories: Practical strategies from the DAEDALUS Project. *Ariadne* (39).
- Max Planck Institute. (2003, 20-22 October). Berlin Declaration on Open Access to Knowledge in the Science and Humanities. Retrieved 28 November, 2007, from <http://www.zim.mpg.de/openaccess-berlin/berlindeclaration.html>
- Nixon, W. (2003). DAEDALUS: Initial experiences with EPrints and DSpace at the University of Glasgow. *Ariadne*(37).
- Open Society Institute. (2002). *Budapest Open Access Initiative*. Budapest, Hungary.
- Pelizzari, E. (2003). Academic staff use, perception and expectations about Open-access archives. A survey of Social Science Sector at Brescia University. Retrieved 24 February, 2005, from http://eprints.rclis.org/archive/00000737/01/Academic_staff_perception_about_Open_archives.htm
- QUT. (2004). Policy F/1.3 E-print repository for research output at QUT. Retrieved 25 March, from http://www.mopp.qut.edu.au/F/F_01_03.jsp
- Rogers, E. M. (2003). *Diffusion of Innovations* (Fifth ed.). New York: The Free Press.

- Sale, A. (2005). The impact of mandatory policies on ETD acquisition. Retrieved 13 Feb 2006, 2006, from <http://eprints.comp.utas.edu.au:81/archive/00000222/>
- Sale, A. (2007). The Patchwork Mandate. *D-Lib Magazine*, 13(1/2).
- Steele, C., Butler, L., & Kingsley, D. (2006). The Publishing Imperative: the pervasive influence of publication metrics. *Learned Publishing*, 19(4), 277-290.
- Suber, P. (2006). Nine questions for hybrid journal programs. *SPARC Open Access Newsletter* (101).
- Swan, A., & Brown, S. (2004). Authors and open access publishing. *Learned Publishing*, 17(3), 219-224.
- Tenopir, C. (2004, 1 February). Online Scholarly Journals: How many? *Library Journal.com*.
- van Westrienen, G., & Lynch, C. A. (2005). Academic Institutional Repositories: Deployment status in 13 nations as of mid 2005. *D-Lib Magazine*, 11(9).
- Ware, M. (2004a). Institutional repositories and scholarly publishing. *Learned Publishing*, 17(2), 115-124.
- Ware, M. (2004b). Universities' own electronic repositories yet to impact on Open Access. *Nature Web Focus: Access to the literature* Retrieved 9 February, 2005, from <http://www.nature.com/nature/focus/accessdebate/4.html>

-
- ⁱ <http://www.openarchives.org/>
- ⁱⁱ <http://www.eprints.org/>
- ⁱⁱⁱ <http://www.dspace.org/>
- ^{iv} <http://arxiv.org>
- ^v <http://repec.org/>
- ^{vi} <http://www.pubmedcentral.nih.gov/>
- ^{vii} <http://www.openoar.org/>
- ^{viii} <http://www.sherpa.ac.uk/romeo.php>