Bond University
ePublications@bond

Library Services and IT Services papers | Library Services

December 2006

Open Source or Off-the-Shelf? Establishing an institutional repository for a small institution

Mark Sutherland
Bond University, mark_sutherland@bond.edu.au

Peta Hopkins
Bond University, peta_hopkins@bond.edu.au

Follow this and additional works at: http://epublications.bond.edu.au/library_pubs

Recommended Citation
Mark Sutherland and Peta Hopkins. (2006) "Open Source or Off-the-Shelf? Establishing an institutional repository for a small institution".Dec. 2006.

http://epublications.bond.edu.au/library_pubs/11

This Conference Paper is brought to you by the Library Services at ePublications@bond. It has been accepted for inclusion in Library Services and IT Services papers by an authorized administrator of ePublications@bond. For more information, please contact Bond University's Repository Coordinator.

Open Source or Off-the-Shelf?

Establishing an institutional repository for a small institution

Mark Sutherland, Associate Director, Information Access Services, Bond University
Peta Hopkins, Information Systems Librarian, Bond University

Abstract

Effective management of digital assets as well as increasing research exposure and impact are particular challenges faced by smaller institutions with limited infrastructure and resources. The paper explores the significant factors involved in considering, planning and establishing an institutional repository for Bond University, one of the smaller higher education providers in Australia. The salient benefits and advantages as well as the disadvantages of implementing an off-the-shelf product as opposed to an open source solution for an institutional repository are compared. The rationale for choosing a proprietary product over an open source solution is discussed, as well as the process for obtaining funding and the support of key stakeholders within the University. The paper describes the strategies employed to populate the repository retrospectively and to train academic staff and researchers in self-archiving. The development of policy governing the repository and intellectual property and copyright implications are also covered.

Background on Bond University

Bond University is Australia's only private, independent, not-for-profit university. By comparison with other Australian universities it is a small institution and one of the youngest universities in the country, soon to be celebrating 18 years since the first cohort of students was welcomed on to the campus in May 1989.

The University offers programs in four faculties: Business, Technology & Sustainable Development; Humanities & Social Sciences; Health Sciences & Medicine; and Law. Bond differentiates itself in Australian higher education by its emphasis on small class sizes, personal attention to students and a three-semester academic year which translates to two-year completion of most of its degrees. Bond is an international university with approximately half of the students and staff coming from overseas. The University maintains a student/faculty ratio of approximately 10:1, the lowest of any Australian University. In 2005 over 4,000 students studied at Bond University, of which approximately 25% were higher degree students. There were 52 higher degree research students in 2005.

Research at Bond University

A key strategic goal of Bond University is "To develop knowledge within and outside the University". This research goal of the University includes objectives to develop and market the University's research profile and to foster research relationships with industry and across the faculties.

In recent years Bond University has qualified for competitive research grant funding and has also been successful in attracting a number of new academic staff actively involved in research. This has added to the number of staff who have been doing excellent research over a long period of time and as a result, research output has increased significantly. This growth in research output is indicated by the number of publications produced.

The following table shows the Commonwealth Department of Education, Science and Training (DEST) reportable research (i.e. collected as part of the Higher Education Research Data Collection process (HERDC):

	2001	**2002**	**2003**	**2004**	**2005**
Books	2	3.5	2.17	4	5
Book chapters	4.76	8.33	14.83	13	16.5
Peer reviewed journal articles	17.61	42.45	48.83	51.53	53.26
Full written conference papers	7.17	15.17	16.17	37.76	35.66
Unweighted publications	31.54	69.45	82	106.29	110.42

Source: AVCC

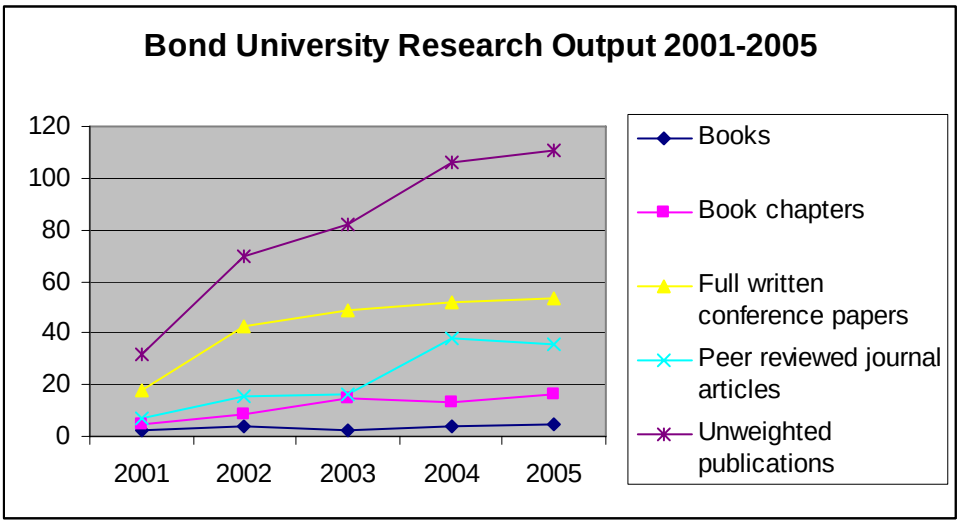

Figure 1: HERDC publications

Early considerations for an institutional repository at Bond

In the climate of increasing assessment of research quality and impact, key academic and Library staff in the University recognised the benefits of an institutional repository to showcase the growing range of research publications at Bond and support the strategic objective of marketing the University's research presence. At the same time the need for effective digital assets management had become apparent as the University embarked on the creation of a digital archive of historical photographs and began investigating ways of creating a digital catalogue of art works on campus.

An embryonic institutional repository using GNU e-prints open source software had already been established as a trial in the School of Information Technology, but it did not have a sustainability plan and housed only a small number of papers. In early 2005 a collection of historical materials of the University was handed over to the Library. This archival collection is large, will continue to grow and contains valuable resources that can be digitised to build a fine heritage collection for the University. The need for

appropriate management of this digital archive was therefore considered in parallel with the need for an institutional repository for research and scholarly literature.

For some time, the University had been keen to explore future publishing options that might include both open access and paid subscriptions to electronic journals published by the University. Therefore the requirements of an electronic press (e-press) were also factored into the investigation of options for an institutional repository.

In addition to research papers, photographs, historical ephemera and Bond-branded journals, theses produced by Bond University students also needed to be taken into account. Submission of electronic versions of theses had been mandated by the University for some time and these were being published on a web server, but by 2005 no progress had been made towards integrating the Bond theses into the Australasian Digital Theses program (ADT) interface. It was therefore envisaged that an institutional repository would effectively manage e-theses as well.

Open source solutions for institutional repositories

Against the background of the paradigm shift in scholarly communication and the increasing move towards open access to scholarly and research literature, as well as the digital asset management issues described above, Bond University Library investigated options for the development of an institutional repository. Through involvement with the Council of Australian University Librarians (CAUL) and the Queensland University Libraries Office of Cooperation (QULOC), Bond University Library was conscious of the development of open source institutional repositories at universities in Australia and New Zealand. These included the repositories set up through the DEST funded initiatives, Australian Research Repositories Online to the World (ARROW) and the Australian Partnership for Sustainable Repositories (APSR).

In particular, Bond University Library was able to draw on the expertise and experience of the University of Queensland (UQ) Library in their establishment of ePrintsUQ (based initially on the open source GNU EPrints2 and later Fedora/Fez). Some of the other open source solutions that were considered included:

- DSpace
- Greenstone Digital Library Software

Proprietary solutions for institutional repositories

In 2004 Bond University Library was aware of a number of proprietary products that could be used for digital content management (mainly learning content management) or digital asset management, but none that were specifically designed for an institutional repository along the lines of some of the open source repository solutions. Some of these proprietary products were:

- Blackboard Content System
- Cumulus (Canto)
- Digitool (Ex Libris)
- Encompass for Digital Collections (Endeavor)
- Hive (Harvest Road)
- Masterfile (Concord Australia)
- The Learning Edge

The ProQuest product, Digital Commons, was a promising newcomer to the scene late in 2004, although initially there were no Australasian customers and the major known repositories were all using open source solutions. However, the product did attract interest as it was designed specifically to accommodate an institutional repository of research and scholarly literature. A pilot project in 2005 aimed at exploring the suitability of Digital Commons in supporting participation in ADT and this highlighted the features and functionality of the product. This was a particularly encouraging development as it was seen as a possible way of including Bond theses in the ADT database through using Digital Commons for the Bond institutional repository.

Open source versus Off-the-shelf

Choosing the best software solution to meet the organisation's mission and objectives is a complex question. While an open source solution superficially appears to be a simple and cost effective alternative, the reality is more complicated. Typically, open source solutions require considerable time, expertise and effort. For a small institution, a resource and staff intensive project with significant requirements was not an option.

Karen Schneider (2006) sums up one factor that was a key consideration for Bond University Library in selecting a repository solution:

> *" …. But I hate the idea that for some librarians if a particular software is open source, hands down, it's the right choice. The right choice is the software that meets the mission. While the principles behind open source are admirable, when an open-source product doesn't meet your library's needs, your first obligation is to your users…..*
> *Software isn't "free" unless the labor to maintain it is "free." Maybe you have the in-house expertise to deal with OSS... but even so, it's still your time and therefore money, and if you don't, you'll have to buy it."*

In a Gartner report, Driver and Weiss (2005, p.19) (Driver and Weiss 2005) conclude that "*Along with potential value, adopters will find higher costs and risks as they target decreasingly mature aspects of the open-source model."* p.19.

As commented in Library Technology Reports, *"… staffing can be the most costly component of an IR project"* (Costs, 2004. p.54). In addition to the staff who will recruit researchers to gain their interest and support as well as upload content, there is a need for technical staff to install, maintain and customise software. For open source solutions, systems administration and programming skills are required for simple tasks such as modifying templates and user interfaces and for more complex tasks where integration with other systems is the goal. Small institutions may find it difficult to budget for or even recruit specialised staff on a temporary basis for these kinds of projects.

Corey Wallis (2006) asked recently, *"… in what circumstances would an organisation choose a commercial product over an open source product, and what benefits do they perceive as gaining by making such a choice?"* The question is specifically related to institutional repository solutions and the following may go some way in answering the question from the perspective of Bond University.

Within the Bond University Library in 2005, the team responsible for IT-related services was fully committed to a range of projects and ongoing service delivery and it would have been very difficult to commit to the additional staffing required to set up an open source based repository.

Not only was the repository project to be absorbed into the existing library staffing budget, tapping into the technical skill set was also going to be a challenge with the University's Technology Services staff committed to the other projects. An open source option was less desirable from that perspective, and it was becoming clear that an externally hosted solution had some real benefits:

- There would be no hardware to purchase, install and maintain – no additional burden on space, air-conditioning or power supply
- It would not be as staff intensive in terms of set up, customisation, configuration and ongoing administration and maintenance
- There would be no delays – uploading of papers could commence immediately after installation
- Backups and redundancy would be the vendor's responsibility

"Redundancy requirements for an IR are likely far greater than the backup currently used for other library systems. While catalogue records can be replaced and a few lost transaction records are not a major calamity, IR contents may be unique and irreplaceable" (Costs, 2004. p.55). This was a major consideration for Bond University and one for which a hosted solution provided the necessary assurances.

A web log post (Hawkins 2006) reporting on a panel session on this topic covers the pros and cons of choosing between open source and proprietary systems. The panel consisted of Jonathan Nabe from the University of Connecticut, a Digital Commons site and Susan Gibbons from the University of Rochester, where DSpace has been so heavily customised that it is apparently difficult to tell that it is actually DSpace – indicative of how much work is actually required to get an open source product to be suitable for a particular institution:

> *"In making the Make or Buy decision, Nabe recommended asking two questions*
> - *Is your technology department under worked?*
> - *Can your internal programming staff match what the vendor can provide?*
>
> *Issues to be considered include backups, archives, and failure protection."*

For Bond University, the answer to both these questions was no, providing clear justification for the choice of an off-the-shelf product.

The following table summarises the various factors that Bond University examined concerning the decision to implement an open source or proprietary solution:

Factors	**Open Source**	**Off-the-Shelf**
Costs	Software is 'free'	Costs cover software licence, support and maintenance
Staffing	Greater need for technical staff for system administration and customisation	Reduced need for technical staff.
Support	Must rely on the community of users for support. Sites may need to be entirely self-reliant.	Support is provided under a contract or agreement. Training is usually available.
Development	Development is often undirected and sites may have divergent needs leading to development in different directions. This is especially true in less mature applications. Risk that extensive customisation will prevent future upgrades of the software.	Development is directed, often in line with input from customer base. Development may be slower, and may not meet all of the individual site's requirements. Enhancements are part of the software development path and will be compatible with upgrades. They may also be automatically rolled out.
Corporate Culture	How does the use of open source software sit within the culture of the organisation?	
Hardware	Hardware for production and development recommended. Set-up and maintenance.	Hardware for production and development recommended. Set-up and maintenance, unless a hosted solution is sought.
Maturity	Immature systems may be more likely to have divergent development paths, and require in-house development, but the 'free' price tag is still a drawcard.	Immature systems are less marketable and if they do not meet requirements are simply not selected.

Figure 2: Factors affecting decisions regarding open source vs. proprietary systems

In the final analysis, the choice of Digital Commons for Bond University's institutional repository was based on:

- The conclusion that an off-the-shelf product was more suitable for the University than an open source solution, based on the above factors
- The fact that Digital Commons would enable Bond's thesis metadata to be harvested for the ADT program as the product is Open Archives Initiative – Protocol Metadata Harvesting (OAI-PMH) compliant
- The attractive pilot pricing of Digital Commons that was offered to Australian universities in late 2005 and the encouraging fact that there appeared to be interest in the product from a number of Australian and New Zealand universities
- The fact that Digital Commons could accommodate images and would therefore meet the University's needs for a digital asset management system, in addition to functioning as a repository for research papers

A firm decision was not made until after ProQuest had demonstrated Digital Commons on the Bond campus. This session was attended by staff from the Library, the Research

Committee, higher degree research students and researchers at the University. Following this demonstration and some further evaluation of Digital Commons sites in North America, as well as obtaining assurances regarding the future migration of content should the University decide to change the repository platform, the decision to purchase an initial 12-month subscription was made.

Obtaining funding and involving key stakeholders

Once the decision to purchase Digital Commons was made in late November 2005, the process for obtaining funding for the project began. As a private, not-for-profit University, Bond is heavily dependant on tuition income and receives very little in the way of government grants. The Library applied to Bond University Research and Consultancy Services (BURCS) for full or partial funding of the subscription for Digital Commons for 2006. Some of the faculties contributed from their research funds and as a result the cost of the product was shared between the research office, some faculties and the Library. In addition to this, the Library offered to provide project staff through budgetary savings and also to provide existing staff to steer and manage the project and provide a reference group during the implementation stage.

The establishment of the institutional repository was a joint venture of the Library and BURCS and frequent updates on the progress of the planning and implementation were discussed at research committee meetings. The Pro Vice-Chancellor, Research also took a keen interest in the project from the outset, as did the various associate deans responsible for research at the faculty level.

To gain the support and interest of academic staff and researchers on campus a presentation by a representative from the University of Queensland about the success of the implementation of UQ e-prints was organised. Titled “Going Global – How institutional repositories enhance your research impact,” the presentation succeeded in building interest in the concept of the Bond institutional repository. In early 2006 an online poll was conducted whereby the entire Bond community was asked to suggest a name for the institutional repository. The winning name was e-publications@bond.

Figure 3: e-publications@bond – home page

Approach to policy development

An initial policy framework based on the Queensland University of Technology (QUT) institutional repository policy was drafted (QUT, 2006). However, unlike other new policies, this policy was not submitted to the University's Quality Task Force straight away. It was felt that much of the initial implementation needed to occur first and that this would inform and guide the evolution of policy, rather than trying to deal with all the issues upfront and having academic staff and others debating the pros and cons without having the benefit of seeing a repository in action. Owing to the fact that there are a number of emotive issues such as intellectual property and copyright involved in establishing institutional repositories, this soft approach to policy development has proved very effective. Ultimately any policy on the institutional repository will have clear links with the University's IP and copyright policies.

Strategies to populate the repository

A project plan for the implementation of e-publications@bond was developed and in it the strategies for populating the repository were detailed. It was envisaged that these would include assisting researchers to transfer their older papers from websites and other sources to the repository and training them in self-archiving. As the project unfolded and initial feedback was received, these strategies were modified accordingly. These strategies, described in more detail below are:

- Using the do-it-for-me model (DIFM)
- Setting up Personal Researcher Pages
- Targeting key academic staff and researchers

The following diagram shows some of the significant milestones in the establishment timeframe which stretched over a period of almost 12 months:

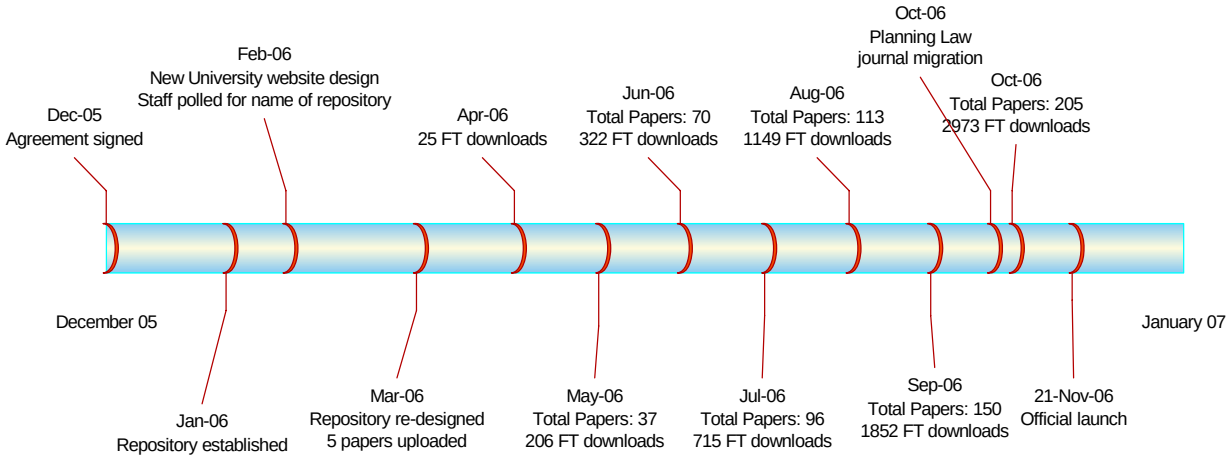

Figure 4: Establishment Timeline – from purchase to official launch

- **DIFM – Do It For Me Model**

Shortly after the initial set up of the shell of the Bond repository, a part-time project officer was hired with the view that much of the work would be to approach academic staff and train them in self-archiving which is supported in e-publications@bond.

However, despite the links for authors to submit papers themselves, there have only been a handful of papers uploaded in this way. It quickly became clear that for most academic staff the DIFM model (do-it-for-me) is much more attractive than self-archiving and this preference was accommodated very early on in the project. However, at no time has the DIY (do-it-yourself) approach been discouraged and the long-term strategy is to train academic staff in self-archiving once their older papers have been uploaded by project staff. The attractions of the DIFM model for academic staff and researchers are:

- The seemingly complex issues of rights management and copyright are handed over to the Library
- The "paperwork" of seeking permissions is handled by the Library unless a publisher specifically asks for the author to request permission
- Digitisation from hard copy, if it is required, is done by the Library and is therefore more consistent across papers in the repository
- Tracking down electronic versions is done by the Library, by staff who are often much more efficient at using article databases and searching the Internet.
- Checking publisher policies is done by the Library

The current strategy is to offer to do the 'legwork' but academic staff are being persuaded to send in their preprints before or at the same time as submitting them to publishers. In part, the aim is to raise awareness among staff about the options for retaining copyright, or at least the right to self-archive in an institutional repository.

- **Personal Researcher Pages**

One of the features of Digital Commons that has not been observed in other repository systems (particularly open source ones) is the facility to showcase research papers for an individual author in one place, in the form of personal researcher pages (PRPs). At Bond, these pages were set up for key staff very early on in the belief that they would help to promote the repository and sell its benefits to researchers on campus. Researchers have found that their PRPs offer an easy way to provide copies of "off prints" with the additional benefit of access to a selection of other documents written by the same researcher.

Not only are academic staff delighted to have their own personal bragging page, but the University's Marketing department has also seen the advantage in the data on these pages and has been working with the Library to investigate options for syndicating this data to staff profile pages on the University website. This is still a work-in-progress.

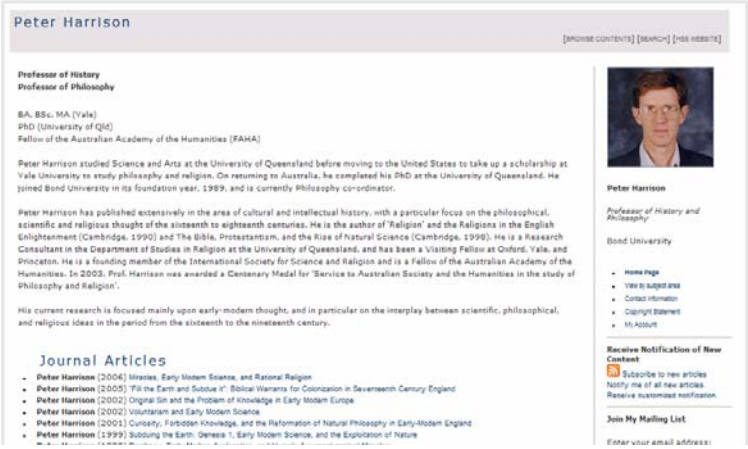

Figure 5: Personal Researcher Page

- **Targeting key academic staff and researchers**

The Personal Researcher Pages also proved to be an excellent way to encourage participation by key academic staff and researchers, particularly those with a higher research profile and or a keen interest in increasing exposure of their research publications.

A Personal Researcher Page was established for each of the faculty representatives in the project sponsor group, the Bond University Research Committee (BURC). Among others, this committee is made up of representatives from each of the Faculties who were in positions that enabled them to spread the word about the institutional repository to their colleagues once a number of their own papers had been uploaded. In addition, academic staff with strong research backgrounds were approached to obtain content for the repository. Other key staff who were sought out included deans and deputy deans of the various faculties.

One of the key selling points of the repository is its role in showcasing the University's research to the world and therefore it was important to get some key papers and names into the repository.

Key research papers such as Steve Webb's groundbreaking paper (Webb, Cupper and Robins) *"Pleistocene human footprints from the Willandra Lakes, south eastern Australia"* were targeted for early inclusion in the repository. Similarly when a Bond academic staff member was appointed to the high-profile Chair of Science and Religion at Oxford University, several of his papers were archived in e-publications@bond and within days a blog post appeared linking directly to the Bond repository.

Keeping up with who and what Bond output is appearing on other websites or in the popular press, and acting to archive relevant papers is an effective strategy in promoting Bond research to the world via the repository. As more links to the repository are posted it has been noticed that Google search results are consistently ranking epublications@bond results in the first 2-3 of the hits. In the early days hits were more likely to appear further down the page or on the second page of results.

While Courant and Griffiths (2006, p.32) in their study of open source software in higher education suggest that faculty support for the aggregation and archiving of research output is not clearly in evidence as yet, this has not been the experience at Bond

University where academics are showing strong support for a central repository of archived research output.

At the time of writing there are over 300 papers that have been archived since the repository was established in March 2006. For a small institution this is considered to be a significant achievement given the time consuming process of obtaining publisher permissions before uploading can commence. Since the first paper was archived there has been a steady stream of requests coming in from researchers to have their papers included in the repository. While there is still a significant amount of content to be added "retrospectively," there is also no evidence that the numbers of submissions will diminish in the near future. In addition, there are long lists of papers for which requests for publisher permissions are pending, or for which pre-print versions are still being tracked down. Almost without exception, academic staff have been willing and enthusiastic to get their papers uploaded.

Internal promotion and marketing of the repository

It was realised early on that an effective strategy for the project was to build the content to a point where the benefits of the repository could be easily demonstrated and marketed. The key researchers in each faculty that were approached for content were nurtured as champions who spread the word. As the repository grew some new features were added including the "most frequently downloaded papers" page, which further helped with marketing initiatives.

The official launch of the Bond institutional repository was held in late November 2006, some 10 months after the implementation began. At that time:

- There were 250 papers uploaded
- A good selection of staff had personal researcher pages
- A two volume digitised book was in the collection
- The newsletter of a research centre had been archived

At the time of the launch there was a significant number of anecdotal stories and evidence that having papers in the repository was advantageous to the author. It quickly became obvious that open access versions of little-used papers were at least being downloaded if not cited, and some authors were actively using the repository to provide copies on request. Several academic staff were very happy to provide testimonials about the benefits to them of having their papers in e-publications@bond. These have been used in a promotional brochure on the repository as well as in a nomination of the project team for the Bond University Vice-Chancellor's Quality Award.

Promoting the repository to the world

To help raise the profile of the repository generally, e-publications@bond was registered with as many repository directories and discovery services as possible. The repository vendor, ProQuest, organised the standard registrations with repository directories as soon as the repository was ready for papers to be uploaded. From the very beginning, Bond papers were being located using Google. e-publications@bond was then registered with the Directory of Open Access Repositories (openDOAR), the Registry of Open Access Repositories (ROAR) and OAIster. After that the repository was registered with the ARROW Discovery Service and later Bond's electronic theses became searchable within the ADT interface.

Registering with repository directories and discovery services seems to work in a similar way to viral marketing. One paper found and linked to in another paper's bibliography or in a weblog would lead readers to find more papers in the same repository. Google is still the chief source of referrals to the Bond repository with most coming from the various regional versions of that search engine.

Unexpected lessons from the project

At the time that this project was conceived, a number of strategies and courses of action were anticipated that in practice turned out quite differently. The flexible and adaptable approach taken in the management of the project allowed this to occur to the benefit of the project as a whole.

For example, from the outset and during the early data-gathering and investigation stage it was expected that an open source solution would be the most likely one, given the fact that most institutional repositories at the time were based on open source software. However, in the end, a fully hosted proprietary product proved to be the ideal solution for an institution such as Bond University.

Based on reports in the literature and on observations made at other institutions, it was anticipated that academic staff and researchers would be trained in self-archiving and assisted to retrospectively add their older research papers. Yet it was found in practice that the do-it-for-me model was preferred by the vast majority of researchers and appeared to work very well.

It was initially felt that all the policy issues relating to the repository needed to be discussed and finalised before the project could properly commence. However, the decision to work with a very basic policy framework and allow the policy to develop and evolve over the implementation period has been an effective strategy.

Finally, concerning publicising and promoting the repository, it was at first thought that a launch with fanfare was required to raise awareness on the campus and ensure the project got off the ground. Yet the soft approach to promotion and obtaining content through involving key individual researchers in each faculty, proved to be a successful way of publicising the repository and ensuring that its vital role was appreciated within the Bond community.

Conclusion

In the Bond experience, establishing an institutional repository for a small institution has been a relatively painless experience and one that has been remarkably successful given the limitations on available infrastructure and resources. This has only been achievable due to the decision to purchase a fully hosted off-the-shelf product. Had the University decided to go down the open source route in this particular instance, one cannot but contemplate that the achievements to date would not have been as significant as they have been.

References

(2004). "COSTS." Library Technology Reports **40**(4): 54-56.

Courant, P. N. and R. J. Griffiths. (2006). "Software and Collaboration in Higher Education: A Study of Open Source Software." Retrieved 21 October 2006, 2006, from http://www.ithaka.org/about-ithaka/announcements/ooss-study-final-report.

Driver, M. and G. J. Weiss (2005). Predicts 2006: The Effects of Open-Source Software on the IT Software Industry. Research, Gartner.

Hawkins, D. (2006). "Institutional Repositories: Make or Buy?" Retrieved 19th November, 2006, from http://www.infotodayblog.com/2006/06/institutional-repositories-make-or-buy.shtml.

QUT. (2006). "Policy F/1.3 E-print repository for research output at QUT." Retrieved 2nd December, 2006, from http://www.mopp.qut.edu.au/F/F_01_03.html.

Schneider, K. G. (2006, July 26). "It's Too Darn Hot: A Curmudgeon's Asides." ALA TechSource Blog, from http://www.techsource.ala.org/blog/2006/07/its-too-darn-hot-a-curmudgeons-asides.html.

Wallis, C. (2006, 29 Sep 2006). "E-Repositories in South Australia." Librariesinteract.info, from http://librariesinteract.info/2006/09/29/e-repositories-in-south-australia/.

Webb, S., M. L. Cupper, et al. (2005). "Pleistocene human footprints from the Willandra Lakes, southeastern Australia " Journal of human evolution **50**(4): 405-413. reproduced in epublications@ Bond -- http://epublications.bond.edu.au/hss_pubs/40/