

Hitting the ground running: Building New Zealand's first publicly available institutional repository

Nigel Stanger, PhD
Department of Information Science

Graham McGregor, PhD
School of Business

University of Otago, Dunedin, New Zealand

Abstract

A fully functional and publicly available, digital institutional repository (IR) in the space of just ten days? The technology was available, the time was right, the team was right and technical assistance from colleagues in Australia was on hand a mere cyber call away. This paper reports on how we were able to “hit the ground running” in building an open access IR in such a short space of time. What has taken our breath away is not so much the speed of the process, but the scale of responsiveness from the Internet community. Consequently, we also consider the research impact of more than 18,000 downloads from eighty countries, less than three months into the project!

Introduction

Digital institutional repositories have become a hot topic over the last two years, and many institutions around the world are now considering or actively working towards implementing them. The University of Otago is no exception, and has implemented a pilot IR within the School of Business.¹

What is so remarkable about the Otago IR? First, its speed of implementation at minimal cost is certainly worth noting, especially for other institutions that are considering an IR. Second, by going live within ten days of inception, it was able to gain recognition as the first public IR in New Zealand. The repository was announced by Professor Arthur Sale (University of Tasmania), a keynote speaker at the *Institutional Repositories Symposium* held by the National Library of New Zealand in Wellington on 23 November 2005. Third, the Otago IR has experienced a spectacular hit rate since the day it went live on 17 November 2005. By 31 January, there had been 9,000 downloads from over sixty countries; just over two months later, the number of downloads has doubled and includes visits from eighty countries.

This article describes how we built a low cost, fully functional repository in such a short timeframe. We reflect on lessons learned and suggest some best practices for implementing an IR. Our reading and discussion with colleagues, especially in Australia, indicated that an IR would enhance the visibility and impact of our research outputs. The level of interest and responsiveness shown by colleagues in the School

¹ We are indebted to our project Research Assistants Monica Ballantine and Jeremy Johnston for their considerable expertise and enthusiasm. Visitors are welcome at <http://eprints.otago.ac.nz/>.

of Business and across the wider university has been most encouraging; particularly in the light of New Zealand's Performance Based Research Funding (PBRF) model. However, we were not prepared for the external volume of traffic that has been generated and are keen to investigate why this should be.

Project background

The New Zealand Government has stated that it is committed to "ensuring New Zealand is a world leader in using information and technology to realise our economic, environmental, social and cultural goals" (New Zealand Government, 2005). They see New Zealand as world leaders in using information and technology to build globally connected science and technology research communities.

In parallel with the launch of New Zealand's Digital Strategy in May 2005, the National Library of New Zealand set up an expert working party with representatives from across the research sector to investigate the feasibility of establishing a national institutional repository for New Zealand's research outputs (Rankin, 2005). The National Library is fostering the creation and launch of a work programme to improve access to New Zealand's research outputs, by collaborating with institutions to stimulate the set-up of research repositories.

Institutional repositories are important because they facilitate greater access to the research outputs from universities, polytechnics, research institutes, other institutions and individuals. The aim is to provide another window on New Zealand research knowledge (Tertiary Education Sector, 2005):

- ensuring open access to the results of New Zealand research
- preserving digital research outputs as an essential record of scholarship
- laying a foundation for exploring new scholarly publishing options
- enhancing and enriching scholarly communication
- joining and contributing to the global IR community

Another driver is the introduction of the PBRF, which was introduced in 2003 and is about to begin its second iteration (Tertiary Education Commission, 2006). The scheme is similar to the UK's Research Assessment Exercise and Australia's proposed Research Quality Framework (RQF), where enhanced research profile has direct funding implications for the tertiary education sector.

The University of Otago's annual publication list ensures that information regarding research outputs is readily accessible, but obtaining the full text of these outputs is more difficult. Unless work is published in a journal that a researcher's institution subscribes to, it is unlikely to be cited (Sale, 2005a; Harnad, 2005; O'Keefe, 2005). Often the most affordable rather than the best work is cited, and Honours, Masters and PhD theses tend to be overlooked as they are less available. An institutional repository provides a centralised location for all publications to be housed or referenced from. The greatest advantage of institutional digital repositories is that they facilitate free sharing of knowledge between colleagues across a range of disciplines at different levels and in widely dispersed locations.

In May 2005, two senior University of Otago staff undertook a study tour of *Digital Challenges* facing universities in the United States. Their report provided the immedi-

ate impetus for the School of Business project. Following discussion with network contacts in the National Library of New Zealand and Australia, a funding application for a proof of concept pilot was proposed to and approved by the School of Business. Project work began on 7 November 2005, with the following goals:

- To establish a proof of concept demonstrator for storing and making available as “open access”, digital research publications in the School of Business.
- To evaluate the potential of the demonstrator for adoption by the wider Otago University research community.
- To connect the School of Business with the global research community, in line with the feasibility study and recommended actions for a national repositories framework for New Zealand’s research outputs (Rankin, 2005).

Building the Repository

The repository was built in three distinct phases: technical implementation, content collection and administration.

Following discussion with colleagues in Australia (we are particularly indebted to Professor Arthur Sale at the University of Tasmania), we used the open source GNU EPrints repository building software, because it was widely used, well-supported, inexpensive and would not lock us into specific technologies or vendors (Sale, 2005b). We also adopted a rapid prototyping methodology, emphasising quick releases of visible results with multiple iterations (Sallis, Tate & MacDonell, 1995, p. 120). This strategy enabled us to create interest in the project at an early stage, and allowed a feedback cycle that would directly involve and benefit stakeholders.

Implementing the software was relatively simple, since GNU EPrints is well established. The only major difficulties that arose were due to poorly-documented platform configuration issues: the first iteration was implemented under Mac OS X, rather than the more typical Linux installation. Most of the time was spent redesigning the look and feel of the web site and tinkering with the internals of EPrints to achieve the desired interface.

Early decisions were made to restrict the content and content domain used for the pilot, in order to speed the collection process and minimise the possibility of project “creep”. In the first instance, content was strictly limited to voluntary contributions in PDF format from colleagues in the School of Business, but with no constraint on the type of output. The project was widely publicised within the School and Heads of Departments were consulted to ensure top-level buy-in. This approach produced immediate results and the repository was quickly populated with a range of working/discussion papers, conference items, journal articles and theses. A sandbox was used to test entries and entry formats before the material went live.

Notwithstanding the attempt to achieve a trouble-free and wide-ranging sample of material in the shortest possible time frame, the need for clear and effective administrative processes proved to be and is paramount. Major issues included:

Copyright: This is a potentially thorny issue for any IR, although many of the concerns raised often turn out to be perceived rather than actual problems (EPrints,

2005). In our case much of the material loaded into the repository comprised departmental working or discussion papers, for which permission to publish online had already been granted. Items with uncertain copyright status had full text access restricted until their status was confirmed. A valuable resource for ascertaining journal copyright agreements is the SHERPA website (2006).

Data standards: New Zealand's Digital Strategy proposes the long term goal of linking all New Zealand repositories to share information and avoid isolated "silos of knowledge", where each institution has little idea of what is happening elsewhere (New Zealand Government, 2005). It is therefore imperative that open standards such as the Dublin Core Metadata Initiative (2006) be applied for both data and metadata. The EPrints software makes this relatively trivial by natively supporting Dublin Core metadata export as specified by the Open Archives Initiative (2006). The University of Otago Library is upgrading to a new catalogue system that also supports Dublin Core, which means that it is possible to directly integrate the repository metadata into the library catalogue.

Data entry: Data entry is likely to be carried out by people who are not specifically trained for the task (for example, document authors), so it is essential to have well-defined and widely publicised processes and standards for data entry. The EPrints software is very helpful in this area, allowing the data entry process to be heavily customised to the needs of an individual repository. In addition, a final verification or editorial step is essential to check the quality of the data entered and to ensure that the item is suitable for inclusion in the repository.

Content acquisition: The key issue regarding acquisition of material is whether self-archiving should be compulsory (top-down) or voluntary (bottom-up). Sale (2005b) argues that a compulsory policy is much more effective at increasing the size of a repository, and illustrates this by comparing the growth rates of repositories at the Queensland University of Technology (compulsory, high growth) and the University of Queensland (voluntary, low growth). Compulsory archiving policies are often driven by the need to capture information for research evaluation and funding purposes, but run the risk that authors may react negatively to such a requirement. Swan and Brown (2004) surveyed 157 authors who did not self-archive and found that 69% of them would willingly deposit their articles in an open repository if required to do so.

Types of content: Decisions about the types of material that should be archived (e.g. working papers, theses, lecture material, sound and picture files) are also key, as is the question of what historical material should be included? There is a cost issue relating to non-digitised work, since scanning or conversion to PDF format is necessary. The value of the repository depends on the number of authors contributing (Rankin, 2005). Ready targets for inclusion are outputs that would otherwise have only limited availability, such as departmental working and discussion papers, and theses and dissertations. The latter in particular are often very difficult to obtain from outside the institution that published them. Paradoxically, however, they are often the easiest to obtain for the purposes of populating an IR, because there is a lower likelihood of copyright issues, and departments often have copies of the documents in question.

The wider University: Given the potential for wider roll out of the pilot, discussion with other key stakeholders was held before the project was formally launched, including the University Librarian and the Director of Information Services. This discussion has created the possibility of a centralised initiative to develop the pilot across the University community. While this possibility is being considered, we have continued to populate the pilot repository. Adding “fresh” items on a regular basis has enabled us to maintain a significant level of interest and downloads, relative to the small number of items (about 220) deposited to date.

How did we get there so quickly?

The Otago IR was fully implemented within ten days of assembling the project team. This outcome was possible because we established a very clear brief to “prove the concept”, rather than taking on a large scale project that would involve many different disciplines, researchers and research outputs from the outset. Meetings were kept to a minimum and policy and procedural issues that required institutional decisions were noted as we progressed, rather than tackled head on. We cannot overemphasise the time and energy that we saved as a result of expert input and advice that we received from the Universities of Tasmania and Melbourne respectively. Apart from their collegiality and willingness to share ideas, there was no associated cost!

Neither of course was there cost associated with the open access software community that we chose to join. From a technical point of view the project was wonderfully straightforward. The Otago IR is deployed on a spare mid-range server running FreeBSD, so our hardware and software costs were essentially nil. In other words, if you happen to have some spare hardware lying around, you can set up an initial repository very cheaply, and then expand it later. The underlying database software for GNU EPrints, MySQL, can scale to very large systems such as the Yahoo! Finance web site, for example (Gilmore, 2001).

We took a minimalist approach to gathering potential content; partly because of the prototypical nature of the project, and partly because material in the hand is worth more than a million promises of what authors suggest they “can” provide if given sufficient time. New publications are always being created, and content acquisition is a moving target that has to be effectively managed. Once some basic content acquisition and data entry protocols were in place, we adopted an incremental methodology. The IR currently contains about 220 documents covering a wide range of topics and document types, and these are added to as new content is acquired. A more systematic approach to content collection is currently being considered.

Finally, it is remarkable what can be achieved by a small, dedicated, knowledgeable and enthusiastic implementation team. As with any project, the right mix of technical and project management skills is crucial in making things happen. Our project team comprised the School’s Research Development Coordinator (project management and evangelism!), an Information Science lecturer (software implementation), the School’s IT manager (hardware and deployment) and two senior students (research, content acquisition and data entry). Oversight was provided by a standing committee made up of representatives from the Information Technology Services Division, the University Library and the School of Business.

Impact of the repository

Has the repository enhanced the School's research visibility and impact? As of 7 March 2006 the statistics show a total of 18,744 downloads from 80 distinct countries since the repository went live. Growth has been consistent over the entire period (see Figure 1). The most popular paper (an Information Science discussion paper) had 451 downloads from 23 countries after twelve weeks in the repository. The top ten downloaded papers included two discussion papers, two working papers, two technical reports and four Honours dissertations (one of which was manually scanned from the original hard copy). Of these, only the discussion papers were previously available online, so the advent of the repository has had a clear impact on the availability of research within the School of Business. Interestingly, the first "officially published" item was a conference paper at position 12; there were 65 conference papers in the repository at the time.

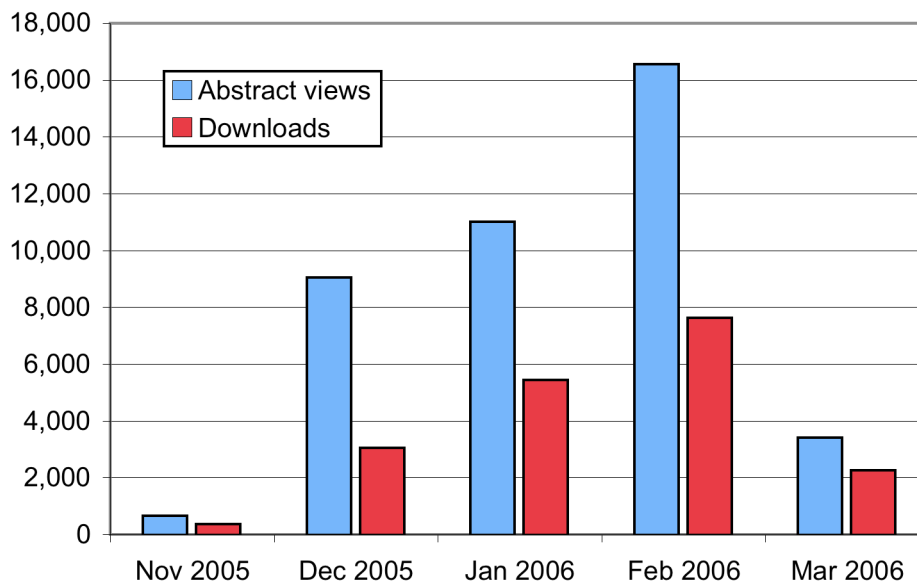


Figure 1. Otago repository download statistics from 17 November 2005 to 7 March 2006.

We have used and extensively modified the statistics package developed at the University of Tasmania (Sale and McGee, 2006) to track downloads of papers from the repository. Rankin (2005) states that through the use of IRs "authors gain visibility, information seekers find research more easily, institutions raise their research profile and funders see wider research dissemination." Research certainly shows a strong correlation between the number of downloads and the number of citations (Antelmann, 2004), and that the citation rates of articles deposited in open access archives increase by 25% to 250% dependent on the discipline (Hajjem, Harnad and Gingras, 2005).

Where to next?

The repository usage statistics indicate unexpected levels of external interest in the Otago IR, which we wish to investigate further. It is unclear *how* the repository become so widely known in such a short time. A possible explanation is a combination

of good timing and early promotion of the Otago IR on relevant web sites, such as the AUSEAccess Wiki (2006), the Registry of Open Access Repositories (2006) and Citebase (2006). Papers from the repository also started appearing in Google search results less than a fortnight after going live. *Why* is it generating so much traffic? Possible reasons could include:

- Otago’s research reputation.
- The reputation of individual authors.
- The first IR to become publicly available in New Zealand.
- The relative paucity of business schools with IRs (the Registry of Open Access Repositories lists only three other business schools, and the OAI lists none).
- Good rankings in search engines like Google (anecdotal evidence shows that the Otago IR generally places well in relevant Google searches).

We look forward to exploring these reasons further through an online questionnaire addressed to repository users during the first quarter of 2006, and would welcome any suggestions or ideas that colleagues may have about how to pursue this work. In the interim, we are also keen to evaluate the impact of the IR within our own institution. We have anecdotal evidence, for example, that the IR has motivated colleagues to complete papers for inclusion in the repository. It has also allowed us to capture historical material that is rare or difficult to obtain. The public nature of the IR has also stimulated colleagues to take interest in the countries that have viewed their papers (see Figure 2); daily checks of download statistics are commonly reported. The statistics and particularly the ability to check for citations are being used as one of the measures for peer esteem as required by New Zealand’s PBRF regime. Given the growing number of requests for demonstrations, the IR appears to be functioning as a catalyst for wider adoption and has created a very valuable feedback loop.

 Finland	38	67
 Thailand	26	55
 Brazil	8	41
 Norway	28	40
 Belgium	54	36
 Latvia	8	34
 Philippines	127	33
 Czech Republic	23	32
 Mexico	23	32
 Switzerland	30	31

Figure 2. A snapshot of the by-country statistics display, showing number of abstract views and downloads, respectively.

Conclusion

This article describes how a prototype IR became fully functional in the space of ten days. The technology has clearly matured to the point where it is neither difficult nor expensive to establish an IR from scratch.

- We have achieved notable results with a core team of five people by setting focussed, achievable and bounded goals.
- The investment in staff time and hardware was small, and there were no software costs. Ongoing costs relating to bandwidth consumption are trivial at NZ\$0.41 per day at peak periods. While these costs may increase as the IR is populated, the concomitant cost/benefit in terms of research visibility and impact appears to be significant.
- Although the repository is small, there is plenty of headroom in the deployed hardware and software, and in any case the repository can be easily migrated to more powerful hardware as necessary.
- Colleagues have been able to make their work immediately available, in a form that can be readily searched by other scholars, and measured in terms of impact through citation tracking.
- The number of downloads and countries that have visited the IR to date is impressive from our point of view. Cliché or not, it has certainly provided our research and researchers with a window to the global world of scholarship.

We recommend setting up an open source IR without reservation, and can only advise those who seek similar outcomes to “hit the ground running” and go for it!

Acknowledgements

Special thanks to Professor Arthur Sale of the University of Tasmania and Eve Young of the University of Melbourne for their enthusiastic assistance and support.

References

Antelman, K. (2004). “Do open access articles have a greater research impact?” *College and Research Libraries* **65**(5): 372–382.

AuseAccess (2006). Wiki site. Accessed on 7 March 2006.

URL: <<http://leven.comp.utas.edu.au/AuseAccess/pmwiki.php>>

Citebase (2006). Web site. Accessed on 7 March 2006.

URL: <<http://www.citebase.org/>>

Dublin Core Metadata Initiative (2006). Web site. Accessed on 7 March 2006.

URL: <<http://www.dublincore.org/>>

eprints.org (2005). *Self-archiving FAQ*. Web page. Accessed on 9 March 2006.

URL: <<http://www.eprints.org/openaccess/self-faq/>>

Gilmore, W. J. (2001). *MySQL powers Yahoo! Finance*, News release, MySQL AB. Accessed on 7 March 2006.

URL: <http://www.mysql.com/news-and-events/success-stories/yahoo_finance.html>

Hajjem, C., Harnad, S. and Gingras, Y. (2005). "Ten-year cross-disciplinary comparison of the growth of open access and how it increases research citation impact." *IEEE Data Engineering Bulletin* **25**(4): 39–46.

Harnad, S. (2005). "Australia is not maximising the return on its research investment." In Steele, C. (ed.), *Proceedings of the 19th Roundtable of the National Scholarly Communications Forum*, 27 September, Sydney, Australia. The Australian Academy of the Humanities.

URL: <<http://eprints.comp.utas.edu.au:81/archive/00000204/>>

New Zealand Government (2005). *The Digital Strategy: Creating our Digital Future*. Policy document. Wellington, New Zealand.

URL: <<http://www.digitalstrategy.govt.nz/>>

O'Keefe, B. (2005). "Rewards await the far-cited." *The Australian*, 19 October, p. 42.

Open Archives Initiative (2006). Web site. Accessed on 7 March 2006.

URL: <<http://www.openarchives.org/>>

Rankin, J. (2005). *Institutional Repositories for the Research Sector*. Feasibility study. Wellington, New Zealand: National Library of New Zealand.

URL: <<http://wiki.tertiary.govt.nz/~InstitutionalRepositories/Main/ReportOfFindings>>

Registry of Open Access Repositories (2006). Web site. Accessed on 7 March 2006.

URL: <<http://archives.eprints.org/>>

Sale, A. (2005a). *Citations*. Wiki article, AUSEAccess. Accessed on 9 March 2006. Last updated 19 November 2005.

URL: <<http://leven.comp.utas.edu.au/AUSEAccess/pmwiki.php?n=General.Citations>>

Sale, A. (2005b). "The key things to know." Presented at the *New Zealand Institutional Repository Workshop*, 23 November, Wellington, New Zealand. National Library of New Zealand.

URL: <http://eprints.comp.utas.edu.au:81/archive/00000223/01/NZ_Workshop_PDF.pdf>

Sale, A. and McGee, C. (2006). *Tasmania Statistics Software*, University of Tasmania. Accessed on 7 March 2006. Last updated 12 February 2006.

URL: <<http://eprints.comp.utas.edu.au:81/archive/00000262/>>

Sallis, P., Tate, G. and MacDonell, S. (1995). *Software Engineering: Practice, Management, Improvement*. Sydney, Australia, Addison-Wesley.

SHERPA (2006). Web site. Accessed on 9 March 2006.

URL: <<http://www.sherpa.ac.uk/>>

Swan, A. and Brown, S. (2004). "Authors and open access publishing." *Learned Publishing* **17**(3): 219–224.

Tertiary Education Commission (2006). *Performance Based Research Fund*. Web site. Accessed on 7 March 2006.

URL: <<http://tec.govt.nz/funding/research/pbrf/pbrf.htm>>

Tertiary Education Sector (2005). *Institutional Repositories: Main*. Wiki article, Tertiary Wiki. Accessed on 7 March 2006. Last updated 6 September 2005.

URL: <<http://wiki.tertiary.govt.nz/~InstitutionalReposities>>