

EPrints makes its mark

Nigel Stanger¹

Graham McGregor²

University of Otago, PO Box 56, Dunedin 9054, New Zealand

¹nstanger@infoscience.otago.ac.nz

²gmcgregor@business.otago.ac.nz

Abstract

Purpose — To report on the impact and cost/benefit of implementing three EPrints digital repositories at the University of Otago, and to encourage others to follow suit.

Design/methodology/approach — Three repositories were successfully implemented at the University of Otago using existing commodity hardware and free open source software. The first pilot repository was implemented within ten days, and is now a fully-functional system that is being championed for institutional-wide use by the University Library. The other two repositories emerged from different community needs. One is academic, concerned with collecting and researching indigenous content; the other is designed to preserve and manage collective memory and heritage content for a small rural community.

Findings — Digital repositories can:

- be established quickly and effectively with surprisingly few resources;
- readily incorporate any kind of extant digital content, or non-digital material that is converted to electronic form;
- meet multifarious needs, from academic institutions seeking to enhance research visibility and impact, to individuals and small communities collecting and preserving their unique memory and heritage records; and
- establish connectivity with the global community from the moment they go live.

Practical implications — The technology and global support community have matured to a state where a fully-featured repository can be quickly and easily implemented.

Originality/value — This article describes the short history, development and impact of the first live repositories of their kind in New Zealand. Their utility and implications for the unique communities that have given rise to them are also explored, by way of encouraging others to take up the digital challenge.

Article Type: Case study

Keyword(s): Digital institutional repositories; Repository implementation; Community repositories; GNU EPrints.

1 Introduction

Digital institutional repositories have become a hot topic in recent years, and many institutions worldwide are now actively implementing them. This article discusses how low cost, yet fully functional digital institutional repositories (IRs), can be set up in a very short time frame. The authors reflect on the lessons learned while implementing three different repositories at the University of Otago, and discuss some new and exciting applications of digital repositories arising from these. The authors also suggest some best practices for implementing an IR and discuss issues that must be considered when moving from a small-scale pilot implementation to a full roll-out.

Interest in institutional repositories at the University of Otago was sparked by the release of the *New Zealand Digital Strategy* by the New Zealand government in May 2005. The strategy aims to ensure that “New Zealand is a world leader in using information and technology to realize our economic, environmental, social and cultural goals” (New Zealand Government, 2005). In parallel with this, 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 a work program to improve access to New Zealand’s research outputs, by collaborating with institutions to stimulate the set-up of research repositories.

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 impetus for the first IR pilot in Otago’s School of Business. Project work began on November 7 2005, with the following goals (Stanger and McGregor, 2006):

- To establish a proof of concept demonstrator for storing and providing open access to digital research publications in the School of Business.
- To evaluate the potential of the demonstrator for adoption by the wider University of Otago 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 (Rankin, 2005).

This article discusses how three different repositories were implemented from scratch, the issues that arose during implementation and the process that has led to their subsequent development and use.

2 EPrints Otago

The GNU EPrints repository management software was chosen for the pilot repository because it was widely used, well-supported, inexpensive and would not lock the School of Business into specific technologies or vendors (Sale, 2005). The development team also had prior experience with the software. A rapid prototyping methodology was adopted, emphasizing quick releases of visible results with multiple iterations, in order to create interest in the project at an early stage, and enable a positive feedback cycle. A sandbox was used to test entries and entry formats before the material went live. Tools, techniques, development tasks and other relevant issues were documented on an ongoing basis using a private wiki.

The pilot implementation was completed within ten days of assembling the project team, with most of this time spent tweaking the look and feel of the web site and collecting content (Stanger and McGregor, 2006). This outcome was made possible by establishing a very clear brief to “prove the concept”, rather than taking on a large scale project involving many different disciplines, researchers and research outputs from the outset. Early decisions were made to restrict the content and content domain for the pilot, in order to speed the collection process and minimize requirements “creep”. Meetings were kept to a minimum and policy and procedural issues that required institutional decisions were noted as work progressed, rather than tackled head on. The project was widely publicized 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.

There was no cost associated with the GNU EPrints software or its associated online community, and from a technical point of view the project was wonderfully straightforward. The School of Business repository¹ was deployed on a spare mid-range server running FreeBSD, which meant that hardware and software costs were essentially nil. In other words, if there happens to be some spare hardware lying around, an initial repository can be set up very cheaply, and expanded later.

A minimalist approach was taken with regard to gathering content; partly because of the prototypical nature of the project, and partly because material in the hand is worth more than promises by authors to supply content at some indeterminate future date. New publications are always being created, and content acquisition is a moving target that has to be effectively managed. Once basic content acquisition and data entry protocols were put in place, an incremental methodology was adopted. Content was limited to voluntary contributions in PDF format from colleagues in the School of Business, but with no constraint on the type of output. As of November 30 2006, the repository contains 409 documents covering a wide range of topics and document types, with new content being continually acquired.

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. The 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 comprising representatives from Information Technology Services, the University Library and the School of Business.

3 Impact of the pilot

Traffic and downloads were generated from the moment the system went live, and the Tasmania statistics package (Sale and McGee, 2006) that sits alongside the repository became an object of fascination in its own right. The initial response to the pilot repository seemed spectacular, with nearly 19,000 downloads recorded within the first three months from eighty different countries. This level of traffic excited considerable interest from both inside and outside the University. However, while the repository had indeed

¹<http://eprints.otago.ac.nz/>

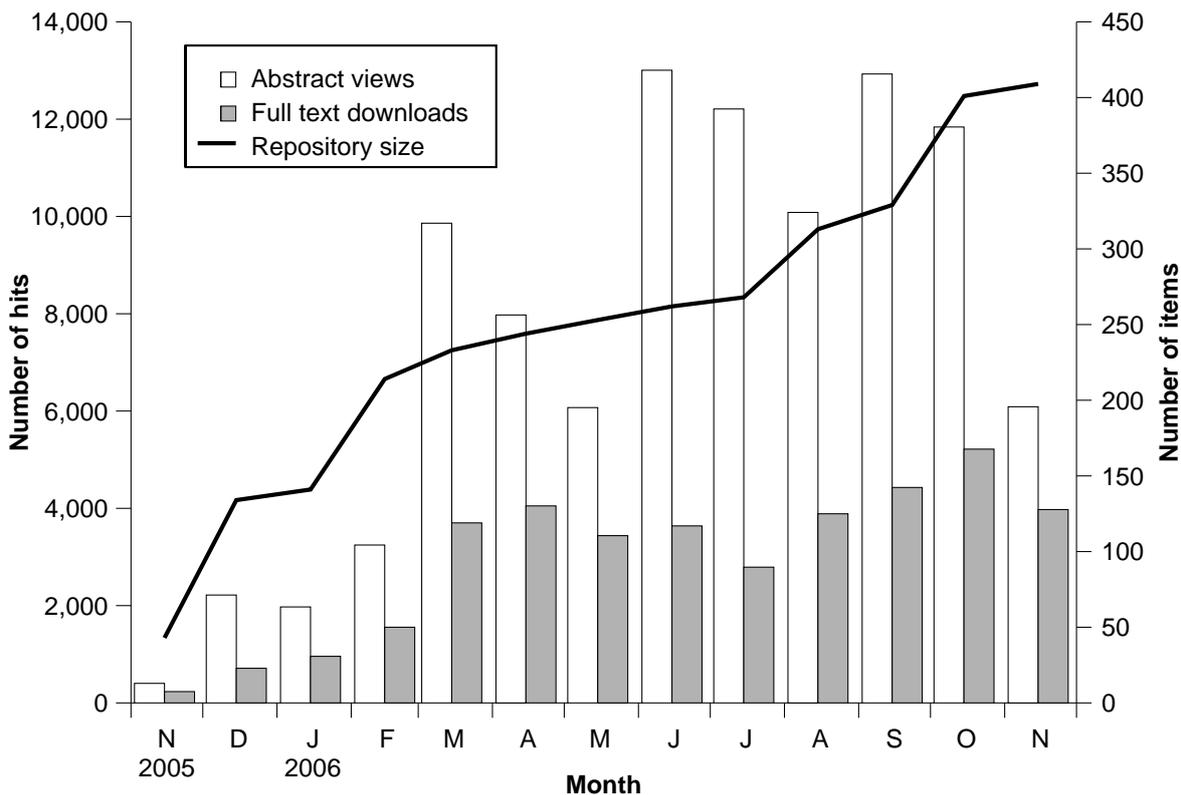


Figure 1: Total monthly hit rates (bar chart, left axis) and number of items (line chart, right axis) for the Otago School of Business repository, up to November 30 2006.

been accessed from eighty countries, it was salutary to discover that the download rates were in fact over-inflated by a factor of about five. This was due to an undocumented assumption in the Tasmania statistics software (Sale and McGee, 2006) that resulted in hits being counted multiple times if statistics were gathered more often than once per day. The lesson here is to always be wary of computers bearing wonderful news!

Despite the downward adjustment to overall download rates, there is still ongoing healthy interest in the repository, as shown in Figure 1. Interestingly, the repository experiences many more abstract views than full text downloads. An informal analysis of hit rates across eight other repositories that generate similar statistics, shows that some experience the same pattern as Otago, while others experience more downloads than abstract views. Further investigation is needed to determine why this variation occurs.

Otago's rate of traffic growth has also been compared with the repositories mentioned above. Figure 2 indicates that traffic to the Otago repository grew much more rapidly during its early months than for any of the other eight repositories investigated, including some that are much older and larger (see Table I). This may be a consequence of growing public awareness of digital repositories, or there may be other factors involved. A research project is currently under way to investigate possible reasons for this finding.

An exciting outcome of the pilot has been the ability to make available material that might otherwise be difficult or impossible to access, and thus increase the likelihood of it being cited (Harnad, 2005; Hajjem, Harnad and Gingras, 2005). For example, Figure 3 shows that nearly three-quarters of the items in the Otago repository are items that might not otherwise be readily accessible, such as theses, dissertations, and departmental

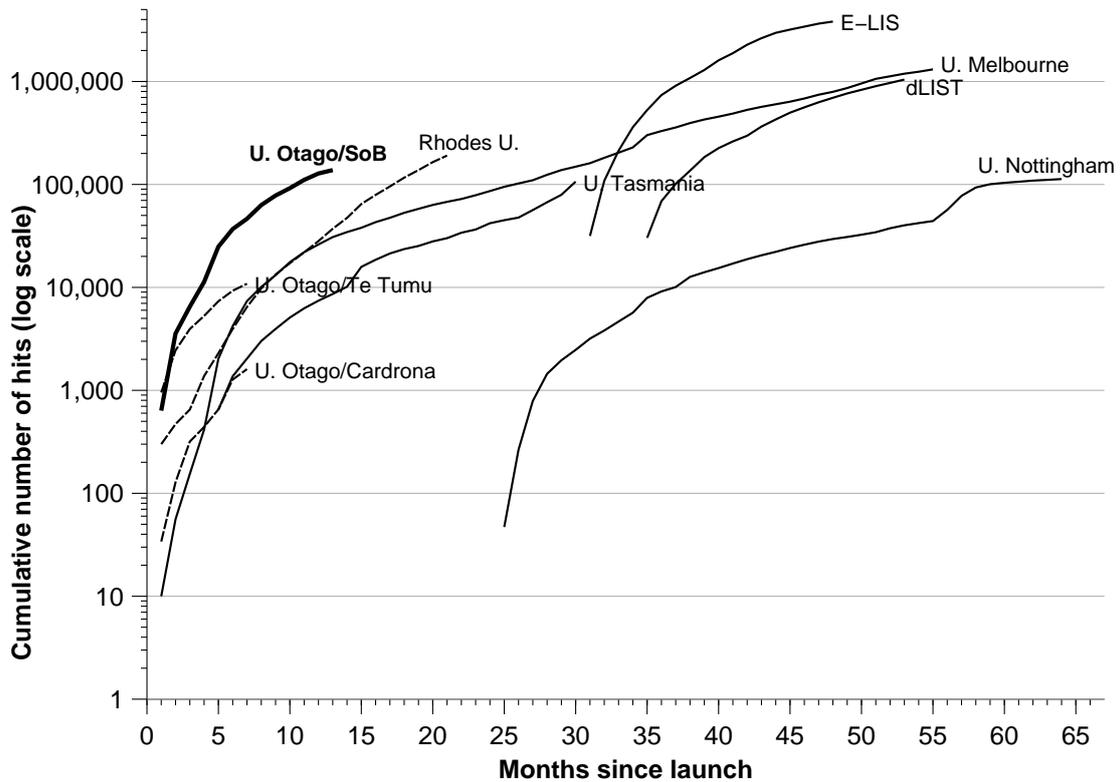


Figure 2: Comparison of traffic growth across nine EPrints repositories, as of November 30 2006. (The different line styles are used only to distinguish the lines; they have no other significance.)

Table I: Details of repositories compared in Figure 2, as of November 30 2006.

Repository	Age in months	Num. items
dLIST (University of Arizona, U.S.A.)	≈ 54	843
E-LIS (CILEA, Italy)	≈ 48	4638
University of Melbourne (Australia)	≈ 53	1479
University of Nottingham (U.K.)	≈ 65	242
University of Otago/Cardrona	6.5	17
University of Otago/School of Business	12.5	409
University of Otago/Te Tumu	7	31
Rhodes University (South Africa)	≈ 21	385
University of Tasmania (Australia)	≈ 26	355

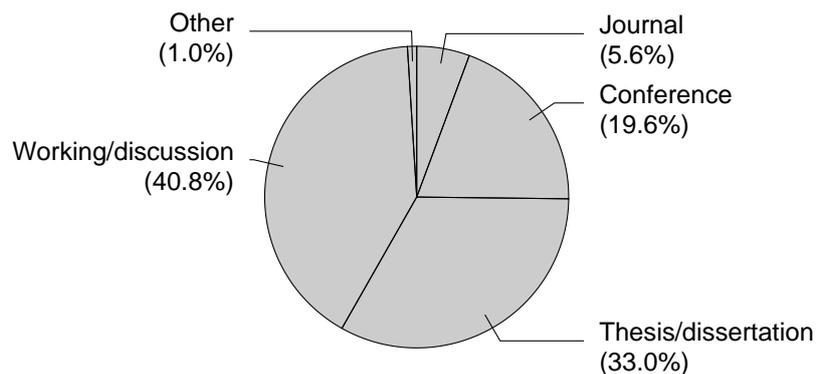


Figure 3: Types of item in the Otago School of Business repository, November 30 2006.

working/discussion papers. Indeed, the top ten downloaded items as of November 30 2006 comprise four departmental working papers, two conference papers, two research reports, one journal paper and one PhD thesis. The full text of these items is also readily searchable by major Internet search engines such as Google (Sale, 2006), often within only a few days of being deposited.

The pilot was not only technologically successful, but also generated much local and national interest. Consequently, after a mere six months, the pilot became the official repository for Otago’s School of Business. It has also been adopted as a model with potential for roll-out across the entire University. As there are four academic Divisions at Otago (of which the School of Business is one), a federated model of repositories is envisaged that would be centrally linked and managed by the University Library.

Having proved the concept, it has been (and is) relatively simple to develop other repositories with similar speed. The key is having an experienced team and a highly focused project management plan.

4 EPrints Te Tumu

The success of the pilot excited considerable interest throughout the University community. In early 2006, Te Tumu, Otago’s School of Māori, Pacific and Indigenous Studies, expressed an interest in implementing a repository for their specific needs. They were particularly interested in the use of a digital repository as a means of disseminating their research and other work, as there are relatively few “official” outlets for their discipline. In addition to the usual items found in most typical IRs, Te Tumu wished to store multimedia items such as images of traditional crafts and artwork, and video clips of performances. This was simply a matter of adding appropriate item types to the EPrints metadata configuration and creating corresponding templates.

Drawing on experience from the pilot, the Te Tumu repository² was implemented in less than a month, and was officially launched on May 3 2006, making it the first repository for indigenous studies in New Zealand. Interest in the repository is evident with almost 4,500 downloads from 70 different countries during its first seven months. The repository currently contains 31 items, including articles, theses, images and video clips.

²<http://eprintstetumu.otago.ac.nz/>

5 Issues to consider

5.1 Copyright

Copyright is an issue that needs to be faced, although concerns that are voiced tend to be perceived rather than actual problems (eprints.org, 2005; Sale, 2006). A substantial fraction of the material loaded into the Otago repositories 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. The SHERPA web site³ was a valuable resource for ascertaining journal copyright agreements.

5.2 Data standards

The *New Zealand 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 Initiative⁴ be applied for both data and metadata. Dublin Core is natively supported by EPrints, and also by many library cataloging systems.

5.3 Data entry

Data entry may often be carried out by people who are not specifically trained for the task (such as document authors), so it is essential to have well-defined and widely publicized processes and standards for data entry. EPrints allows the data entry process to be heavily customized to the needs of an individual repository. A final editorial verification is also essential to check the quality of the data entered and to ensure that the item is suitable for inclusion in the repository.

5.4 Content acquisition

The key issue regarding acquisition of material is whether self-archiving should be compulsory (top-down) or voluntary (bottom-up). Sale (2005; 2006) argues that a compulsory policy is much more effective for growing a repository, as illustrated by 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 access repository if required to do so. A more recent study increased this figure to 81% (Swan, 2006).

Another issue is when authors should deposit new content into a repository. In particular, should pre-prints of submitted papers be immediately deposited, or should authors wait until the paper has been accepted for publication? There are valid arguments for both positions, but in the case of highly popular repositories, waiting for acceptance may

³<http://www.sherpa.ac.uk/>

⁴<http://www.dublincore.org/>

prove to be a “safer” option. In March 2006, the authors submitted an article to a journal, and concurrently deposited a pre-print (Stanger and McGregor, 2006) into the pilot repository. The pre-print quickly became the most popular download from the repository, with 625 downloads in only three weeks. The journal subsequently rejected the article on the basis that the material had already been widely disseminated and was therefore no longer topical.

5.5 Types of content

Decisions about the types of material that should be archived (e.g., working papers, theses, lecture material, multimedia files) are also key, as is the question of what historical material to include. Indeed, this has proved to be one of the most challenging issues faced at Otago, since there can be considerable cost associated with scanning to convert non-digitized work into digital format. There are also associated practical and logistical issues.

The value of a 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, yet paradoxically, 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 to hand.

Extant and already bound material requires page-by-page scanning, which can be a long and arduous process. While a number of robotic scanners are available, these are likely to be out of the financial reach of most institutions. The content focus at Otago has thus moved towards the development of a mandatory policy that requires all student theses and dissertations to be submitted in both hard and electronic copy.

6 Looking ahead

An exciting consequence of the School of Business repository has been an approach from various communities throughout New Zealand to help set up repositories of heritage material relating to their community. The first of these was Cardrona, a small rural Central Otago community with a long and varied history. The Cardrona Community Repository⁵ was launched on May 17 2006, and is the first community repository in New Zealand. Digital repositories offer communities a wonderful opportunity to preserve their historical and cultural heritage, and to disseminate it to a much wider audience than normally possible. It can also provide a sense of focus for the community, especially in cases like Cardrona, where the population is quite small and somewhat geographically dispersed. This information can be of academic use too, such as in a recent study that used community historical information to document the long-term effects of climate change (Hopkin, 2006; Miller-Rushing, Primack, Primack and Mukunda, 2006).

The Otago team is also playing a significant role in the Open Access Repositories in New Zealand (OARiNZ) project⁶. This is a government-funded project to develop a national infrastructure connecting all of New Zealand’s digital research repositories.

⁵<http://cardrona.eprints.otago.ac.nz/>

⁶<http://www.oarinz.ac.nz/>

Work is currently under way at Otago on an easy-to-use installer and configurator for EPrints repositories, in order to encourage wider adoption of these technologies.

7 Conclusion

Experience at Otago demonstrates that in an increasingly digital world, digital repositories are a necessary and welcome means of archiving and making accessible electronic content of all kinds. Global connectedness between scholars and communities at the touch of a keyboard is not a clichéd dream, but a reality. The technology has matured to the point where a basic repository can be set up with a very moderate level of technical expertise. Even setting up a heavily customized repository can be achieved in a matter of days rather than weeks, if a dedicated and knowledgeable team is created and given focused, achievable and bounded goals. Software costs are essentially nil, hardware costs are minimal, and there is a hugely supportive and generous worldwide community of scholars who are willing to share their technical knowledge and expertise at no cost.

On the non-technical side, there are now sufficient repository implementations around the world that IR's are becoming less of a novelty and more an integral tool for researchers, librarians and archivists alike. While Otago is yet to adopt an institution-wide repository, there is little doubt that the progress made to date with its three different thrusts has generated widespread interest locally, nationally and globally. In a purely academic context, the tension between traditional (journal based) scholarship and publishing, and digital (repository based) scholarship and publishing has yet to play itself out.

The authors' experience with community preservation and heritage groups, on the other hand, suggests that given appropriate access to the technology, the content flood gates will truly open. The imprint of EPrints at Otago has not only made its mark, it has stimulated a renaissance-like enthusiasm for making available knowledge and ideas and history and scholarship that might otherwise remain hidden or inaccessible. The added value is that the required institutional or community investment, both time and money, in developing a digital repository seems rather trivial. The authors suggest that prospective repository developers "hit the ground running" and welcome contact from anyone who needs help to do so!

Acknowledgments

The authors would like to thank Professor Arthur Sale of the University of Tasmania, Eve Young of the University of Melbourne and Stevan Harnad of the University of Southampton for their enthusiastic assistance and support. The authors are also indebted to project Research Assistants Monica Ballantine and Jeremy Johnston for their considerable expertise and enthusiasm, and to School IT Manager Brent Jones for deploying and maintaining the repository server. A final acknowledgement must go to Te Tumu and the Cardrona community for the wonderful opportunities that they have provided.

References

eprints.org (2005), 'Self-archiving FAQ', web page. Accessed on September 26 2006.

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

- 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* **28**(4), 39–46.
- Harnad, S. (2005), Australia is not maximising the return on its research investment, *in* C. Steele, ed., ‘19th Roundtable of the National Scholarly Communications Forum’, The Australian Academy of the Humanities, Sydney, Australia.
*<http://eprints.utas.edu.au/204/>
- Hopkin, M. (2006), ‘Family albums highlight climate change’, web article, *news@nature.com*. Accessed on 27 September 2006.
*<http://www.nature.com/news/2006/060807/full/060807-10.html>
- Miller-Rushing, A. J., Primack, R. B., Primack, D. and Mukunda, S. (2006), ‘Photographs and herbarium specimens as tools to document phenological changes in response to global warming’, *American Journal of Botany* **93**(11), 1667–1674.
- New Zealand Government (2005), ‘The Digital Strategy: Creating our Digital Future’, policy document, New Zealand Government.
*<http://www.digitalstrategy.govt.nz/>
- Rankin, J. (2005), ‘Institutional Repositories for the Research Sector’, feasibility study, National Library of New Zealand.
*<http://wiki.tertiary.govt.nz/~InstitutionalRepositories/Main/ReportOfFindings>
- Sale, A. (2005), ‘The key things to know’, presented at the New Zealand Institutional Repository Workshop, Wellington, New Zealand.
*<http://eprints.utas.edu.au/223/>
- Sale, A. (2006), Researchers and institutional repositories, *in* N. Jacobs, ed., ‘Open Access: Key Strategic, Technical and Economic Aspects’, Chandos Publishing, Oxford, UK, chapter 9, pp. 87–100.
*<http://eprints.utas.edu.au/257/>
- Sale, A. and McGee, C. (2006), ‘Tasmania Statistics Software’, University of Tasmania. Accessed on September 26 2006.
*<http://eprints.utas.edu.au/262/>
- Stanger, N. and McGregor, G. (2006), Hitting the ground running: Building New Zealand’s first publicly available institutional repository, Discussion Paper 2006/07, Department of Information Science, University of Otago, Dunedin, New Zealand.
*<http://eprints.otago.ac.nz/274/>
- Swan, A. (2006), The culture of Open Access: researchers’ views and responses, *in* N. Jacobs, ed., ‘Open Access: Key Strategic, Technical and Economic Aspects’, Chandos Publishing, Oxford, UK, chapter 7, pp. 52–59.
*<http://eprints.ecs.soton.ac.uk/12428/>
- Swan, A. and Brown, S. (2004), ‘Authors and open access publishing’, *Learned Publishing* **17**(3), 219–224.

About the authors

Dr. Nigel Stanger is a lecturer in the Department of Information Science at the University of Otago School of Business, where he has taught in the areas of systems analysis and database systems since 1989. He has active research interests in digital repositories, distributed and web database systems, XML technologies, physical database design and database performance. He was the project lead and programmer for the School of Business EPrints repository, which he continues to maintain and enhance. He is also heavily involved in projects to increase the uptake of digital repository technology within New Zealand, and is a key member of the Open Access Repositories in New Zealand (OARiNZ) project.

Dr. Graham McGregor is the Research Development Coordinator for the University of Otago School of Business. He is an experienced tertiary academic and manager, who has held senior positions in both the polytechnic and university sectors in New Zealand, and worked as an independent consultant. As an academic, he largely published in the field of sociolinguistics. He has also joint authored work on ICT pedagogy and practice and has written reports for several New Zealand government agencies. His current role is to stimulate and coordinate research development activities across New Zealand's business and academic communities. He was instrumental in launching the School of Business EPrints repository.