



Research Repository Case Studies

For presentation and discussion at the “Research Repository Managers Symposium”

EDUCAUSE Australasia 2009, Perth Western Australia. 3-6 May 2009

<http://www.caudit.edu.au/educauseaustralasia09/> Session/Paper number 99.00

Prepared by Convenors

1. Leonie Hayes - University of Auckland, Auckland, New Zealand l.hayes@auckland.ac.nz
2. Teula Morgan - Swinburne University of Technology, Melbourne Australia tmorgan@swin.edu.au
3. Tom Ruthven - University New South Wales, Sydney, Australia t.ruthven@unsw.edu.au

Link to Case Study <http://hdl.handle.net/2292/3368>



This work is licensed under a [Creative Commons Attribution 3.0 New Zealand License](https://creativecommons.org/licenses/by/3.0/nz/).

Conference Abstract:

A Research Repository Managers Symposium invites managers to submit a “Case Study” outlining the way that their institution has decided to deliver the requirements for ERA – Excellence in Research for Australia and PBRF Performance-Based Research Fund in New Zealand. The symposium session asks authors of the case studies to briefly share their case studies, followed by a guided discussion session determined by participants. The Case Studies will be compiled into a comprehensive document for public distribution via the Educause Australasia 2009 Conference site.

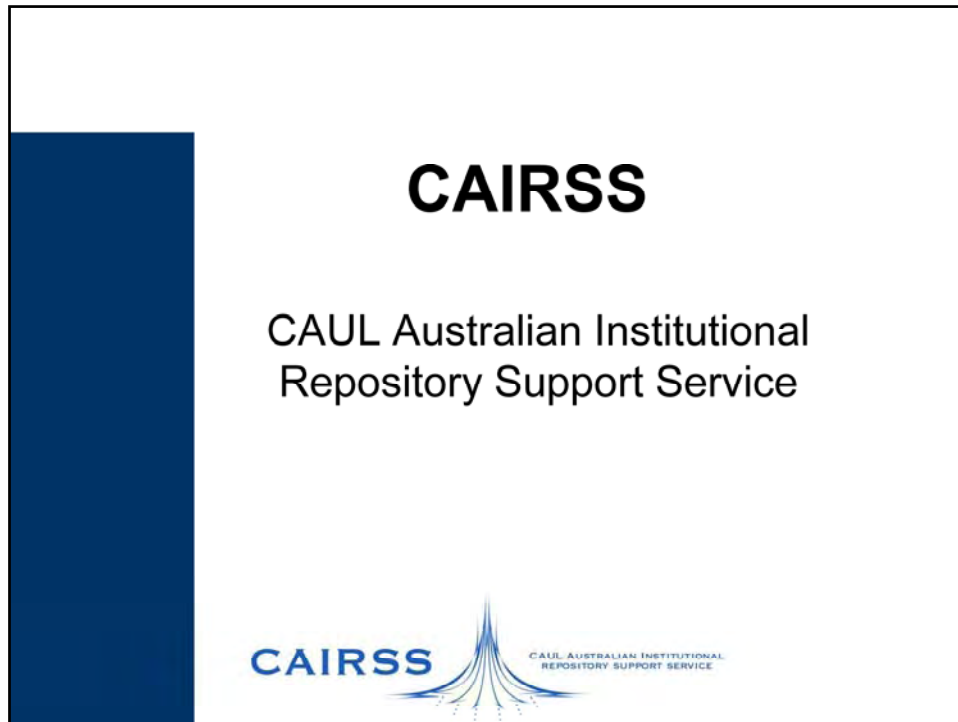
For other similar Case Studies see the ones compiled by Open Repositories 2008 Conference in the UK <http://pubs.or08.ecs.soton.ac.uk/86/> The focus of this symposium is how Research Repositories support tertiary institutions in delivering Research Data Collection in Australia and New Zealand.

Themes and information to address in the Case Study would be

- Institutional overview
- Models – comment on your institutional model based on the Arrow HERDC report
- <http://www.arrow.edu.au/docs/files/arrow-herdc-workinggroupreport.pdf>
- Institutional Embedding and Innovative Practices
- Relationships with Research Management Systems
- Sustainability
- Outreach, Marketing and Faculty Engagement
- Technical Environment and Information Technology strategies
- Lessons learned and Future Plans.

Contents

1.	CAIRSS – CAUL Australian Institutional Repository Support Service	Australia
2.	Central Queensland University	Australia
3.	Charles Sturt University	Australia
4.	Curtin University	Australia
5.	Macquarie University	Australia
6.	Murdoch University	Australia
7.	Swinburne University of Technology	Australia
8.	University of New South Wales	Australia
9.	University of South Australia	Australia
10.	University of the Sunshine Coast	Australia
11.	University of Sydney	Australia
12.	University of Western Sydney	Australia
13.	Victoria University of Melbourne	Australia
14.	Auckland University of Technology	New Zealand
15.	Canterbury University	New Zealand
16.	Lincoln University	New Zealand
17.	Massey University	New Zealand
18.	University of Auckland	New Zealand
19.	Victoria University of Wellington	New Zealand
20.	Waikato Institute of Technology	New Zealand
21.	Waikato University	New Zealand
22.	Cape Peninsula University of Technology	South Africa
23.	Council for Scientific and Industrial Research	South Africa



Introduction

CAIRSS (CAUL Australian Institutional Repository Support Service) has been running for seven weeks now. Nearing the end of implementation phase, the service is funded to run for two years.

Following discussions with the Department of Innovation (DIISR) the Council of Australian University Librarians (CAUL) Executive, and the ARROW Management Committee, CAUL established the CAUL Australian Institutional Repository Support Service (CAIRSS), to provide support for all institutional repositories in Australian universities, regardless of the software being used. CAUL appointed the University of Southern Queensland, through the Australian Digital Futures Institute (ADFI) within the Division of Academic Information Services (DAIS) to undertake its new institutional repository support service.

To offer support for Repository Managers
in the higher education sector in Australia.

The service aims to:

1. Provide a forum to represent the collective interests of repository managers around Australia
2. Support and develop toolkits
3. Provide best practice and policy advice
4. Assist with the integration of repositories with the requirements of the ERA HERDC exercises
5. Assist the understanding of managing copyright issues in the repository environment
6. Provide a watching brief on trends and developments in repositories



CAIRSS Aims

The primary function of CAIRSS is to offer support for Repository Managers in the higher education sector in Australia.

The service aims to:

1. Provide a forum to represent the collective interests of repository managers around Australia; support emerging areas of activity;
2. Support and develop toolkits for copyright and institutional repositories;
3. Provide best practice and policy advice for areas such as data migration, metadata, standards compliance, import and export, harvesting, ingest of new forms of digital material;
4. Assist with the integration of repositories with the requirements of the ERA and the Higher Education Research Data Collection (HERDC) exercises;
5. Assist the understanding of managing copyright issues in the repository environment;
6. Provide a watching brief on trends and developments in repositories.



CAIRSS Central

Similar to the Repository Manager community which it supports, CAIRSS Central staff are located all over Australia:

- Project Manager and CAIRSS Repository Specialist, Katy Watson (the key contact for the service) is located at University of the Sunshine Coast in Maroochydore, Queensland. Katy has worked as a repository manager since 2006 and prior to taking this position was working across the Library and Office of Research at USC implementing research systems for integration with institutional repository activity.
- CAIRSS Technical Officer Tim McCallum is the CAIRSS contact for all things technical. Tim has previously been an integral member of the successful RUBRIC Central team. Tim is based at University of Southern Queensland in Toowoomba, Queensland.
- Luke Padgett is the CAIRSS Copyright Officer and Swinburne University of Technology Copyright Advisor. Located at Swinburne's Hawthorn campus in Victoria Luke manages copyright services delivered to Swinburne, the CAIRSS community and through a contract of service the University of Ballarat.
- Dr Peter Sefton is the CAIRSS Senior Advisor.
- And for those who do not know me, I am Professor Alan Smith, the USQ delegate for CAIRSS.

Behind the service is CAUL's Steering Committee led by Helen Livingston of the University of South Australia, with Judy Stokker of Queensland University of Technology, Heather Gordon of James Cook University, David Groenewegen of ANDS, the Australian National Data Service and CAUL's Diane Costello.

Interacting With CAIRSS

Informal Communication

- Twitter – use the #caulcairss tag
- CAIRSS Blog – <http://caulcairss.wordpress.com/>
- Instant Messaging – details coming soon

Semi-formal Communication

- CAIRSS Discussion elist <http://groups.google.com.au/group/cairss>
- CAIRSS Topic Teleconferences
- CAIRSS Wiki

Formal Communication

- Email CAIRSS – cairss@caul.edu.au
- Phone CAIRSS - 07 5430 2823 or 0419658448
- Submit a CAIRSS task trac ticket
- CAIRSS website - cairss.caul.edu.au

See News Stand page on the CAIRSS website
http://cairss.caul.edu.au/www/news_stand.html



Interacting With CAIRSS

CAIRSS will facilitate interaction amongst Repository Managers and Repository Technical Officers in the higher education sector in Australia. CAIRSS Central hopes to create an interaction model with various communication tools and different levels of communication to support this community. It is hoped the CAIRSS community grows organically with the repository community.

Informal Communication

CAIRSS Central would like create an opportunity for the community to share issues and ideas in a relatively informal manner if they wish, allowing for open discussion where repository managers and technical staff at different stages can interact.

This is where issues and ideas can be communicated casually, enabling community members to ask any question regardless of topic, experience or expertise. This will be useful to get the ball rolling on topics of interest to the community and may result in semi-formal communication where these issues and ideas can be further investigated by CAIRSS.

- Twitter – use the #caulcairss hash tag to monitor and contribute to CAIRSS tweets
- CAIRSS Blog – <http://caulcairss.wordpress.com/>
- Instant Messaging – details coming soon

Semi-formal Communication

This is where issues and ideas can be discussed collaboratively by the CAIRSS community and CAIRSS Central. This semi-formal level of communication will involve asking questions to the community, sharing of ideas, experimenting, and testing with the end result being collaborative solutions.

- CAIRSS Discussion elist – Join at: <http://groups.google.com.au/group/cairss>. Post to: cairss@googlegroups.com. This elist currently has 85 members.
- CAIRSS Topic Teleconferences – These will be held regularly. The first was held on 29 April on the topic 'Including Creative Works in your Repository for the purpose of ERA'
- CAIRSS Wiki – details coming soon

Formal Communication

This is where issues and ideas can be submitted directly to CAIRSS Central for processing.

- Email CAIRSS – cairss@caul.edu.au
- Phone CAIRSS - 07 5430 2823 or 0419658448
- Submit a CAIRSS task trac ticket – details coming soon

Formal communication will also include the CAIRSS website (<http://cairss.caul.edu.au/>). Guides resulting from work undertaken by CAIRSS Central as requested by the CAIRSS community will be added to the CAIRSS website.

Please note: You can subscribe to RSS feeds for all of these tools on the [CAIRSS News Stand page](http://cairss.caul.edu.au/www/news_stand.html) on the CAIRSS website http://cairss.caul.edu.au/www/news_stand.html, or alternatively call Katy Watson and she will happily introduce you and your staff to each communication tool.

Next Month

- Phone hook-ups with each CAUL Institutional Repository Manager to ensure CAIRSS Central has a clear picture of each institutions repository progress/issues, and also to discover what each institution needs/wants from the CAIRSS Service (to help facilitate organic growth of the service)
- Providing login details for using the CAIRSS Wiki and CAIRSS Trac system for logging jobs
- Phone hook-ups with each CAUL Institutional Repository Technical Officer to ensure they have access to the current suite of repository technical tools, and also to discover what other technical tools they require to assist in the running of their repositories
- Working with repository technical staff to provide guidance on configuring each repository system for ERA authenticated access
- AURC – Australian University Repository Census
- Future topic teleconferences



Next Month

Planned CAIRSS Central activity for the next month to support Australian Institutional Repository Managers and Repository Technical Staff include:

- CAIRSS Project Manager Katy Watson having phone hook-ups with each CAUL Institutional Repository Manager to ensure CAIRSS Central has a clear picture of each institutions repository progress/issues, and also to discover what each institution needs/wants from the CAIRSS Service (to help facilitate organic growth of the service)
- CAIRSS Technical Officer Tim McCallum calling each CAIRSS community member providing them with their login details for using the CAIRSS Wiki and CAIRSS Trac system for logging jobs
- CAIRSS Technical Officer Tim McCallum having phone hook-ups with each CAUL Institutional Repository Technical Officer to ensure they have access to the current suite of repository technical tools, and also to discover what other technical tools they require to assist in the running of their repositories
- CAIRSS Technical Officer Tim McCallum working with repository technical staff to provide guidance on configuring each repository system for ERA authenticated access
- AURC – Australian University Repository Census
- Future topic teleconferences

EDUCAUSE 2009

Research Repository Case Study

CQUniversity Australia (aka Central Queensland University)

Introduction

The University's repository , ACQUIRE (www.acquire.cqu.edu.au) was established in 2006 as a joint venture between the Division of Library Services and the Office of Research to capture and store the University's research output. Since then it has been used to capture and report the University's annual HERDC research outputs.

In 2007 the ACQUIRE team received a Vice Chancellor's Award for Excellence in recognition of services to the University community.

We are now tackling the challenges of ERA.

Institutional Overview

CQUniversity Australia is a small regional university based in Rockhampton, Queensland with campuses in Mackay, Emerald, Gladstone and Bundaberg and Brisbane, Gold Coast Sydney and Melbourne. We also have a large cohort of students studying via distance education.

Details regarding the campuses and courses on offer are available at our website - <http://www.cqu.edu.au/>

The repository is managed by the Division of Library Services in collaboration with the University's Office of Research and Information Technology Division. ACQUIRe is a digital repository designed to enhance Central Queensland University's research output by:

- Providing free web access to publications by CQU staff and students, where copyright permits
- Increasing visibility of CQUniversity's research publications through open access compliant records that can be harvested by major search engines, including Google
- Increasing impact and citation rates of CQUniversity's research publications
- Facilitating collaboration between researchers by enabling them to easily share data and publications
- Preserving CQUniversity's research output for the future

The repository was originally branded aCQUIRe - *a Central Queensland University Institutional Repository* - and it has been formally registered as a University trademark. It was a good idea at the time but the lower case at the start and end of the acronym caused havoc with spell-checkers etc. whenever staff were referring to it in correspondence or other documents. Consequently, earlier this year, we changed the acronym to all caps – ACQUIRE to alleviate this problem.

As part of the establishment process for the repository an Institutional Repository Policy (http://policy.cqu.edu.au/Policy/policy_file.do?policyid=679) was endorsed by the University's Academic Board.

HERDC Model

Submission to the repository for HERDC publications was mandated shortly after the repository's creation and replaced the print-based submission forms previously used by the Office of Research.

At this stage print copies of publications can still be forwarded to the Office for verification but we are planning to mandate the submission of electronic version in the near future.

There are 3 review stages within the submission process where various staff in the Library and Office of Research check, amend and add information as required before the researcher(s) submission is added to the repository. These stages are as follows:

- ACQUIRE Eligibility Review where Liaison Librarians are allocated their discipline-specific submissions to validate that the research publication submitted is
 - eligible according to the policy and
 - check the auto-created citation
- Office of Research Review where the HERDC eligibility of the publication is assessed and verified by staff in this area.
- Copyright Review where, as its name suggests, the publication's copyright status in regards to whether it can be publicly displayed is assessed using OAKList; sherpa , publisher web sites etc. and requests are made of publishers as required.

Outreach, Marketing and Faculty Engagement

We have just adopted the '*NISO-RP-8-2008 - Journal Article Versions (JAV): Recommendations of the NISO/ALPSP JAV Technical Working Group, April 2008*' as a way of reducing the confusion surrounding the terminology of pre-prints and post-prints. The recommended terminology of Author's Original (or AO) = preprint; Accepted Manuscript (or AM) = postprint and Version of Record (or VoR) = publisher's pdf or publisher's HTML or publisher's XML, should go a little way towards making submitting into the repository less complex for researchers and librarians alike.

In conjunction with the Office of Research, our Research Liaison Librarians regularly conduct HERDC information and drop-in sessions for academic staff to help them with their submissions.

At this stage we are still waiting for statistics to be fully and reliably enabled so that researchers can start seeing the fuller benefit of having their research output in ACQUIRE.

Liaison Librarians also promotes the repository to new and current academic staff during their Faculty / School visits and at relevant meetings and forums.

The repository has also been actively promoted by several University hosted conferences and authors advised that their papers will be submitted to the repository as a condition of acceptance.

We are also marketing ACQUIRE with branded USB sticks and stress balls.

Relationships with Research Management Systems and other systems

Since 2006 the reporting functionality of the Library's Information Management System (LIMS) has been utilized to form the basis of the University's HERDC submission to DEST / DIISR.

The University has recently purchased ResearchMaster and the Office of Research is progressively implementing the various modules.

It is on our agenda to develop other interfaces to the repository such as output to personal web pages, CV creation, bibliographies and social networking tools, but the impact of ERA has meant that these are on the backburner for the time being.

Research Assessment and Reporting

As mentioned above the repository data has formed the foundation of the University's HERDC report. since 2006.

As mentioned before the lack of accurate statistics is a barrier to progressing with research assessment and reporting on a wider scale.

Technical Environment and Information Technology strategies - Platforms/Architectures

- We are using VTLS' Vital (repository) and Vital (online submission) software solutions and were a member of the ARROW consortia.
- We are using the expertise of existing cataloguing staff and the cataloguing module of the Library's Information Management System (LIMS) for metadata cleansing, author authority control and the addition of at least one LCSH heading.
- Once submitted, the MARCXML datastream created in the repository is extracted, converted to MARC and imported to the Library's cataloguing module for the abovementioned functions.
- Once completed, the MARC record is saved to the catalogue so the information for all the University's research publications is available when staff and students search for resources. The catalogue records have a hyperlink to the repository where access is then available to the full text (copyright permitting). Our link resolver provides access options for journal articles not publicly available.
- A copy of the final marc record is then also extracted, converted to MARCXML, loaded back in to the repository and the Dublin Core regenerated to reflect the changes made to the data in the catalogue.
- ACQUIRE is being harvested by the ARROW Discovery Service (<http://search.arrow.edu.au/>).

Sustainability

At this stage we are still in project mode with the majority of staff funded from the University's ASHER grant. This year we are concentrating on ensuring that all the necessary research publications are in the repository to support the University's ERA 2009 trial and 2010 submissions.

The project implementation of the repository and associated government initiatives (RQF and now ERA) are managed through an ACQUIRE Working Party which encompasses representatives from the Library, Office of Research and Information Technology Division.

With our strategy to use the catalogue for the storage and maintenance of our data, we would rate the quality of our metadata as excellent. Our main issue at the moment, is keeping the data between the catalogue and the repository synchronised.

The ratio of Open Access full text versus hidden repository content is about 50/50 and this basically revolves around copyright and the fact that researchers, at this stage, very rarely keep their manuscript versions once published.

Lessons learned and Future Plans

What is next?

- 1 -2 years
 - Capturing and describing creative works for ERA.
 - Data synchronization
 - Researcher engagement and acceptance through accurate statistics, author pages etc.
 - Links to/with ResearchMaster.
 - Revisiting our software solution - in-house solution versus hosted solution
- Long-term
 - Increased depositing of research outputs as a result of government research grant mandates.
 - Increased acceptance by academics and recognition of open access benefits
 - Increased full-text availability of author(s) manuscript versions.

What has been easy to solve?

- As the whole process has been, and continues to be, such a steep learning curve for everyone involved I'm not sure we can say anything has been easy to solve. We have done the best we can at the time and are therefore continually revisiting decisions made as our knowledge increases and the developments surrounding repositories continue to evolve.

What has been difficult to solve?

- Academic engagement and recognition of repository and open access benefits
- Non-availability of author manuscripts.
- What advice would you give others?
 - Make sure you have a strong relationship with your Research Office
 - Ensure you have good IT support.
 - Identify and utilise key research champions to populate and promote the repository.

Elke Dawson,
Deputy Director
Resource and Access Services
e.dawson@cgu.edu.au
07 4930 6838
8th April, 2009

Repository Case Study for Educause

Name of Institution: Charles Sturt University (CSU)

Repository Name: CRO CSU Research Output <http://researchoutput.csu.edu.au> 1

Authors: Karin Smith, Sarah Bishop

Summary/Abstract: CSU has developed CRO, an integrated repository and research management system, to provide access to the University's research output in an open access environment and to meet research reporting requirements such as HERDC. The repository was used for the first time in 2008 to report CSU's 2007 HERDC submission and as a result some significant changes to the database structure have been required. Further development work is concentrating on refining CRO for the ERA Trial and full ERA exercise in 2010.

Date: 30 March 2009

2. Institutional Overview

Charles Sturt University is an inland regional university with campuses spread over central and southern NSW. The University's vision is to be: *A national university for excellence in education for the professions, strategic and applied research and flexible delivery of learning and teaching.* The CSU Vision and Mission statement are available at: <http://www.csu.edu.au/about/mission.htm>

CSU is the largest provider of distance education in Australia. The total enrolled number of students in 2007 (latest available statistics) was 33,926 of which 21,284 were studying by distance. Total student FTE for the same year was 18,367. CSU's research strengths cover a variety of disciplines and foci including: philosophy and ethics, theology, environment, water, viticulture and wine production, agriculture, rural sociology, education, professional practice, health sciences

The repository is running under project methodology which has expanded to cover the ERA requirements. It is managed by the Library's Development Section as it is dealing with new and evolving digital information environments. There is a very close collaboration with the CSU research office, the Centre for Research and Graduate Training (CRGT). Repository administrative duties are shared by the Library and CRGT. Repository hosting and technical support is provided by UNILINC Limited (which also supplies the Library's LMS). UNILINC works closely with CSU to develop and customise the repository software, DigiTool, to meet our needs.

The repository is named CRO an acronym of CSU Research Output. The name was chosen by the Steering Committee in early 2007 and has now become widely synonymous with the repository at CSU. It fits within similar naming conventions at other institutions such as University of Wollongong's RO or Griffith University's GRO. Advice to those yet to choose a name would be to avoid using a product name in the title as this can be restrictive if you migrate to a different repository software platform at a later date. The same applies to a choice of domain name.

The Project was scoped and defined in the later quarter of 2006, with the proviso that it would be hosted externally to the University. The repository went live in December 2007 when it received its first HERDC submissions. The forms and original data dictionary were developed in the four months before that. During the next six months all CSU HERDC data was collected in CRO for 2007. This data was reported to DEEWR from CRO. Lessons learned from the first HERDC exercise using CRO indicated that the repository's data structure needed to be revised to achieve more refined reporting. Since then development work has involved re-evaluating the Dublin Core (DC) fields (DigiTool default) and creating a process of converting the data to MARC21 which allows for more reliable reporting. CSU participated in the ERA Pilot exercise in October-November 2008 which

interrupted the data conversion work but proved a worthwhile exercise in light of the forthcoming ERA Trial.

The CRO Submission policy was passed by Academic Senate in October 2007 and outlines the mission of the repository:

The CSU Research Output (CRO) Repository presents publicly available research and scholarly outputs. It contributes to the growing international trend for Open Access to scholarly outputs and makes research literature freely available on line. Material submitted to the repository must comply with the Submission Policy to be made publicly accessible. The full text material can not be added to the Open Access Repository if copyright restrictions exist. In such cases citation and abstract will be added. Embargoes will be observed and material can be added to CRO when the embargo is over.

The policy mandates the submission of prepublication versions of eligible items. This has been a key principle from the beginning and was further strengthened in October 2008 by requiring all research accepted for publication to be added to CRO at time of acceptance.

- *All refereed revised and final draft research manuscripts are required to be deposited in the CSU Institutional Repository (CRO) [immediately] after they have been accepted for publication except for books which may be self-archived at the author's discretion.*
- *Such manuscripts be required to be made Open Access available to anyone on the web except where this is restricted by a publisher's policy.*
- *ALL CSU staff are required to submit their peer-reviewed pre-publication manuscripts for the DEST Higher Education Research Data Collection (HERDC) as an electronic file into CRO from 2007 onwards.*

The Policy is available at: <http://www.csu.edu.au/research/publications/cro/policy.htm>

3. Comment

CRO has been developed to be a combined repository and research management system for HERDC data collection and reporting. In this respect it conforms most closely to Model 4 in the ARROW HERDC Report: a combined repository and research management system. The requirement to combine functions was defined at the beginning of the project to address the fact that CSU did not have a separate research management system. CRO development was predicated on the understanding that the repository would be the primary instrument for gathering HERDC data for reporting purposes and to capture open access compliant versions of documents for exposure.

By aligning submission to CRO with the annual HERDC process it has been relatively easy to incorporate the repository into institutional practice. Researchers were already used to entering HERDC submissions into web forms, the CRO submission forms are an improvement on the old process which also helped our cause.

One of the key requirements of the repository is that information is only updated on CRO which is the master data source for CSU research outputs and associated metadata. Shadow systems are not maintained ensuring reports are always generated from the current information. To achieve this we have had to retrofit our data structure (noted above) to ensure the reporting is always accurate as there had been some data slippage in the old DC structure. A standard institutional repository requires a much simpler data structure, for which DC is perfect. The additional research management element has requirements that DC cannot meet because of the lack of subfields in its structure. At time of writing the data conversion work is still not complete for some of the records in

CRO. This is an issue as further development work is dependent on its completion. It is worth noting that what seemed like an obvious combination – combining the repository and research management system has led to very interesting challenges and if anyone is thinking of doing it they may like to speak with us. Our implementation was very rushed because we wanted to report from HERDC right from the start and have needed to 'retrofit' as a result of the learning experience.

4. Outreach, Marketing and Faculty Engagement

Marketing and promotion of CRO is embedded into the existing processes of the institution because it is tied to HERDC and part of this is the mandated process of adding prepublication versions of articles to the submissions. Faculty are engaged in contributing to the repository by necessity. This is underpinned by timely directives from the DVC-Research (DVCR) and by payment of funding incentives based on research points. High-level support from the DVCR, deans and other senior executives of the University will be further leveraged for ERA.

Workflow: records submitted to HERDC are verified and quality assured by the library and CRGT. CRGT checks data accuracy and eligibility for HERDC, the library checks copyright and processes and badges the files. Files remain locked and unprocessed until this checking is completed. This system is evolving. After the first year's HERDC submission process, forms and workflows were re-evaluated and amended for 2008 HERDC submission. Changes were also made to the submission forms to match the change in CRO's data structure.

Copyright: The submission process requires authors to acknowledge copyright and licence the repository to perpetually hold the item and apply any necessary preservation measures.

OA rights are determined through checks in SHERPA/Oaklist databases. If the information is not available then the publisher is asked whether the item can be loaded. Publisher responses are kept as a record and their instructions are followed. There is some disquiet among academics about being required to submit their prepublication versions however most do it as it is mandated and tied to a financial incentive.

5. Relationships with Research Management Systems and other systems

As noted in Point 3 CRO is a combined repository and research management system as far as HERDC and other research output reporting is concerned. CRO does not contain any data related to research income, this information is captured and stored externally to the repository. However ERA has signalled the need to capture and store additional new data and/or report data stored in other systems. Data sharing linkages will be needed with the human resources system and the University's finance management system. CRO does not provide the facility to host personal web pages at this stage however the CRO Policy requires researchers to link citations on their personal webpages to the full text in CRO. Improved reporting should also enable the generation of bibliographies for CVs etc.

6. Research Assessment and Reporting

The repository was designed to capture relevant metadata and digital objects and to generate reports to fulfil HERDC requirements. It was always planned to build on this structure for the RQF and latterly, ERA. To that end, in addition to capturing 2007 + HERDC submissions, historical metadata from 2002-2006 HERDC reports was loaded into the repository in 2008. This data was used to create CSU's submission to the ERA pilot in October 2008. The repository will be used for all research output data for ERA trial and in ERA in 2010.

To date we have not developed any useful metrics information from CRO. Further developments for CRO are dependent on the finalisation of the data structure conversion from DC to MARC21 (which

we hope will be completed shortly). Once complete it is planned to register CRO for OAI harvesting and concentrate on developing usage reports. Feedback we've received from CSU researchers indicate that these are important tools for them and will be an additional incentive to archive their work to CRO.

7. Technical Environment and Information Technology strategies -Platforms/Architectures

Software Platforms, Database and operating Systems: DigiTool – the ExLibris proprietary institutional repository product with technical support provided by UNILINC Limited.

Versions:

Software: DigiTool v3.2

Database: Oracle v9.0

Operating System: RedHat Linux

The system is hosted externally at UNILINC Limited and backed up nightly. There is also a test system used for development and which can be used for submission in the event of significant downtime in the production system. There is no mirror site.

Customisation: CRO extends the capabilities of the DigiTool software rather than extensively customising it. This approach ensures that development of the repository does not stray from the basic software design so CRO can handle DigiTool updates and new versions when released. The significant change to the database structure is possible because DigiTool supports both Dublin Core and MARDC21. DigiTool submission forms use Dublin Core fields. We have some unique local fields, but have used DC schema where possible. We have crosswalked that to MARC21 and the submitted data (will be) converted to MARC21 in the repository – this enables more precise reporting bases on the MARC fields. Data is converted back to DC for display, indexing and harvesting.

There has been some minor customisation of the web interface such as headers in the search and submission screens. This customisation requires adjustment each time a service pack update is issued.

We have not yet set up to harvest to OAI as the system is still being developed. CRO is not registered for open URL linking although this may be a later development since CSU also uses ExLibris's SFX and Metalib products.

8. Sustainability

Staffing and Resources: The repository is running as a project, extended for ERA, managed by the Library. The project manager and the repository manager are Library staff members. CRO is the responsibility of the Division of Library Services but, as noted above, has close links and shared work processes with the research office. The Research office staff work on the repository but do so as part of the HERDC and ERA work. There is no dedicated position in the research office for ongoing repository work, other than HERDC. Responsibility for managing and resourcing CRO will revert to the Library at the end of 2009. The Library plans to employ a part time clerical assistant from 2010 to provide ongoing assistance to the repository manager. Additional assistance for ERA preparation will be resourced from project funds. As repository development is principally focussed on meeting ERA requirements ASHER funding is being utilised.

Management structure: The repository manager reports to the Director Development, Division of Library Services. While the repository remains under project conditions, the Director Development is also the project manager. A steering committee convened by the Executive Director, Library Services with membership of DVC Research and a nominee of the Executive Director Information Technology provide governance of the project.

Quality of metadata: all records are submitted by academics. The 2007 HERDC data needed some editing. The earlier, 2002-2006 data was surprisingly accurate considering it was also self submitted. The research centre officer did not spend a great deal of time editing it. All historical data will need to be checked again for ERA. Each HERDC record needs to be checked when submitted, edits are almost always made. CRO does not use subject headings or other controlled language. This was a deliberate decision as we consider the researchers will use the most appropriate terms when describing their research, and because we expect most discovery will be via OAI searching across full text records rather than specific visits to the CRO search interface.

Open Access Full text: This was very high for 2007 HERDC submissions as it was a mandatory condition. About half of the 1000 records submitted for 2007 HERDC are available for open access. The rest either cannot be released or weren't sent. We are happy to load the prepublication version and do not seek out permission to load the publisher version. CRO has the facility to lock non-compliant documents and to manage embargoes.

The 2002-2006 data does not have articles attached. Content will be sought as part of the ERA data collection process but the mandate to submit full text for all records won't be enforced for his historical data other than where required for ERA assessment.

Open access compliance hasn't been checked for 2008 HERDC data yet so statistics are not known although most submissions contain the prepublication version in accordance with mandated policy.

9. Lessons learned and Future Plans

Planning: Most of the coming year will focus on developing the repository for ERA including recruiting additional data and digital objects. Creative works forms have been created and will shortly be trialed with targeted researchers. Feedback will be used to refine the process and identify and gaps.

Having started later than some, CSU does not have a digital theses collection. Development of this collection in the repository is also planned but will be a secondary priority to ERA.

Longer term plans are to investigate how the repository will support eResearch and possible support for an ePress. This timeframe will be governed by University priorities and external developments such as the ANDS initiative.

Advice: If you are starting from the point of developing a combined repository and research management system, focus on planning what is required of the research management system and a metadata structure to support it. As long as your chosen software supports this metadata and the storage of digital objects, the repository will follow out of that.

Developing such a system is IT resource-hungry. Ensure you have adequate and dedicated resources. No repository software (open source or proprietary) that we are aware of has been designed to address the separate requirements of research management and institutional repository. Developing a hybrid requires time and dedication.

A visionary, engaged DVC-Research who could see the value of combining the two systems and mandating the submission of articles has been invaluable in the success of CRO.

10. Additional Information

More detail and background on the development of CRO can be found in the following paper.

Bishop, S., Smith, K., Cargnelutti, T., King, B. (2009) Pushing the envelope: making a good system great through collaboration. Retrieved 30 March 2009, from http://www.information-online.com.au/sb_clients/iog/bin/iog_programme_C8.cfm?vm_key=561C49BB-F5FD-2B70-1DBB26A859E3ED3E

- 1. Please give us the following details • Name of Institution • Name of the repository(ies) and web address(s) • Author(s) of the case study (include contact details if possible) • Summary/abstract • Date prepared**

Curtin University of Technology Espace@Curtin <http://espace.library.curtin.edu.au/>

Julie Woodland Repositories Librarian - Curtin University Library Tel: 08 9266 4203 Fax: 08 9266 4185 Email: j.woodland@curtin.edu.au

Prepared : March 2009

Curtin University Library was the first Western Australian University Library to “take the plunge” into the repository field in 2004 – having an operational repository by 2005. In a pre-RQF, pre-ERA environment we charted a lot of new water, encouraging academics and researchers to participate and expose their research to the world in yet another digital environment. The lessons learned along the way have proved useful in preparing the Library, and the University, for the current research environment.

- 2. Institutional Overview Include information about your institution, size, focus, FTE, culture • Where the repository is managed from • Partnerships, collaborations • Name of the repository – any tips for others on choosing a name • Repository timeline • Mission statements • Regulatory environment within your institution**

Curtin University of Technology is Western Australia’s largest university. In 2007, the university had a total of 41,348 students and 2,890 staff. Education and research is divided across five areas: The Centre for Aboriginal Studies, Humanities, Science and Engineering, Health Sciences and the Curtin Business School.

The University promotes itself as valuing cultural diversity, ethics and equality and conducts programmes in regional and overseas campuses. The University Library manages the institutional repository. The Library collaborates with the University’s Office of Research and Development in relation to sharing of publications data and developments in ERA collection and reporting.

Repository name: espace@Curtin Tip – It is a good idea to incorporate flexibility and association with the institution into the name. The name of our repository is espace@Curtin – this name was selected to indicate the “e” or “digital” nature of the content, and branded it with a Curtin identity. By not including the word “eprints”, the repository would be ready to evolve into accepting other digital formats or data, without the need to change its name.

The repository has been operating since 2005, with software evaluation, policy formulation, configuration and consultation occurring during 2004. The public search page for the repository describes it as follows: Welcome to espace@Curtin, Curtin's research repository. Espace is an open access digital collection containing the research output of Curtin staff and students. Included are digital theses, journal articles, conference papers, book chapters and more. espace preserves Curtin's research output and increases its exposure and impact by making it visible and freely available through open access services such as Google Scholar

and ARROW (Australian Research Repositories Online to the World).

The Library is responsible for the planning and management of the repository. The library is also able to consult with the University's Research committees on repository policy matters if relevant. Consultation also occurs with relevant groups e.g. Legal and Compliance Services where this is useful to clarify or interpret issues.

3. Comment on • Your institutional model based on the Arrow HERDC report • <http://www.arrow.edu.au/docs/files/arrow-herdc-interimreport-june08.pdf> • Institutional embedding and innovative practices

espace@Curtin most closely resembles Model 2: Research management system to institutional repository. Process described as follows: Data is manually entered into the Research Management System (Script) by academics or associated staff. Records are automatically "tagged" for overnight ingest into the repository (espace@Curtin).

The repository performs a direct database query on the Research Management System. Library staff process records before making "live" and viewable via the repository interface. The URLs of new records in espace@Curtin are fed back to the Script system. That is, each Script record will display a URL linking to its related repository record. Curtin University Library was the first library in Western Australian to implement an institutional repository. It was one of the first Australian libraries to design and implement a data-sharing mechanism between its repository and the University publication system in 2005.

The library was able to work with the Office of Research and Development as they replaced their old publications system, and build in functionality that populated the repository with entries to the University's Research Management System. The "integrator" system helped to build repository content, embed the concept of repositories into the University' research culture and prepare for further developments in publications reporting.

4. Outreach, Marketing and Faculty Engagement Give details about any of the following • Strategies and Services • Support required • Workflow • Frequently asked questions • Copyright, licencing,(include use of Creative Commons) and legal aspects • Barriers to moving forward

Most early involvement of academics was by individual approach from the repository manager and faculty librarians who were well known to researchers. Presentations were made at relevant staff functions and meetings as the question-and-answer approach did much to alleviate concerns in the neo-repository era. A small team of trained library staff processed repository records amongst their other duties.

In 2008 ASHER funding has allowed two dedicated (including 1 casual) staff to focus on this task following migration of the repository to a new platform (Digitool). Library staff verify, edit and check each record's metadata as imported from the Script publications system. Copyright restrictions on each record are checked before creating access privileges on any document.

The required attributions and links are attached to each record according to the publishers' requirements. Contributors (i.e. Curtin authors) complete a form agreeing to use of the material. Any legal issues can be clarified with the University's legal office – however the majority of decisions regarding the treatment of material can be made via resources such as Sherpa/Romeo, publisher's websites or via direct request to publisher's/learned societies.

The library will design a formal communication plan to augment the current promotion carried out by repository staff and Faculty Librarians. This plan is likely to use the current emphasis on research assessment to promote the repository as an integral part of research output management for all relevant Curtin staff and research centres.

5. Relationships with Research Management Systems and other systems • Interoperability, data exchanges, and ongoing relationships • Data duplication and/or matching • Single signon and authentication issues • Output to personal web pages, CV creation, Bibliographies and Social Networking tools

espace@Curtin articulates with the University's Research management System. Data is manually entered into the Research Management System (Script) by academics or associated staff. Records are automatically "tagged" for overnight ingest into the repository - (espace@Curtin). The repository performs a direct database query on the Research Management System.

Library staff process records before making "live" and viewable via the repository interface. The URLs of new records in espace@Curtin are fed back to the Script system. That is, each Script record will display a URL linking to its related repository record. Records originally imported from the Script system were ingested with an entry date of 2005 or later.

The system cumulates daily, avoiding the routine ingest of duplicates. Duplication can be checked if the error has been entered on the Script system twice. Use of the Research Management System requires normal University authentication.

There is no direct deposit function into the repository, eliminating the need for contributors to authenticate. The URLs of new records in espace@Curtin are fed back to the Research Management System (Script). That is, each Script record will display a URL linking to its related repository record. Academics or faculties may use these URLs for any purpose. For example, the Curtin Business School routinely harvests these URLs from Script to populate staff publications web pages. For the purposes of HERDC reporting or ERA, this espace@Curtin link in the Script system will provide a view of the research output for evaluation purposes.

6. Research Assessment and Reporting Comment on • How your repository supports formal research assessment exercises • Other reporting exercises generally • Discus on metrics etc outside of these formal exercises

The repository uses the records already entered in the University's publications management system i.e. already entered for internal assessment/promotion exercises and HERDC reporting. The repository will support the ERA trial in 2009 by presenting the top

20% of University research output in a repository environment as required by the Government. The Library will be working with the Office of Research and Development to develop a set of records which meet ERA specifications. Academic staff and research students use the repository to track downloads of their abstracts and text of their document. The system records both types of access, and presents this information to the author in a statistical table. They are able to check rates of access to their material, in addition to how many downloads were made of their colleagues' material.

**7. Technical Environment and Information Technology strategies -
Platforms/Architectures • Software Platform(s) • Database and Operating Systems •
Unique setups • Mirrored Systems, Redundancy and backup strategies •
Customisations – how heavily is your system customised? • Metadata schemas • Open
URL linking • Harvesting**

The repository uses Digitool software, an Ex Libris product. Then data was migrated from the original Eprints software to Digitool in November 2008. The Digitool application is running on a Solaris 10 cluster. The database is Oracle 10g hosted on a Linux cluster. The application is on a cluster and is in the process of being configured to fail over to an alternate node should a problem occur. The Oracle database is hosted on the University's centrally managed Oracle cluster.

Other than the standard configuration of the Digitool 3.0 software for our own purposes we have heavily customized elements of the public user interface. We have also added a number of bespoke statistics gathering and reporting mechanisms. Metadata is Dublin core. Data is harvested via OAI-PMH protocols.

8. Sustainability • Organisational Structure - Staffing and resources • Funding and external grants • Management structure – reporting lines • Rate the quality of your metadata • How much Open Access full text does your repository contain? – is it a question of resources or copyright?

In 2009, a Repositories Librarian position oversees the management and development of the system and service. The position coordinates the training and work of repository support staff. There are two full time positions at Library Technician level processing records into the system, assisted by 4 – 6 other staff at Librarian level, contributing time as available during the "quieter" periods of semester. Funding provided by ASHER grants has employed a full time Library technician casually from January – May 2009.

The Repository Librarian reports to the Manager, John Curtin Prime Ministerial Library, who has overall responsibility for the development of digital archives and special collections. Metadata is checked and verified with the published version of a document on ingest from the Script publications system. Library staff edit citations to conform to a Chicago-style format for consistency. Documents' access rights are described one of five ways.

This allows for variations in the ways digital objects can be treated according to permissions. Open access – an estimated 65% of the repository's content is open access. That is, a visitor to the site is able to gain access to the full-text/content of the digital object attached to the

record's metadata. Where required or useful, the metadata includes links to content on external sites.

Copyright and publishers policies influence the ability to make the remainder open access.

9. Lessons learned and Future Plans • What is next? 1 year? 2 years? Longterm • What has been easy to solve? • What has been difficult to solve? • What advice would you give others

The repository has good coverage of all Faculties in the University. However, emphasis will be placed on the continued growth of content and systematic capturing of all research centres and groups. Current collection policies will be reviewed to determine whether there is scope to broaden the types of material included and whether there is the demand or opportunity to extend the repository's scope.

The repository's future directions will be guided by potential developments in the University's research or data management plans. Most technical issues have been easily and quickly resolved with in-house technical support by staff that both understand the nature and intent of the repository, and have provided customisation since its inception in 2004.

The system on both platforms required extensive configuration but has provided good functionality with minimal maintenance. In the establishment period, collection and usage policies were guided by formal consultative processes with University research committees. Whilst these processes occupied the time of senior research staff, they provided a forum to provide discussion and ratify the repository's proposed directions. It is important to have a communication process in place with stakeholders to ensure that the repository is aligned with the University's priorities for research support.

Macquarie University Library

Macquarie University ResearchOnline – a digital repository case study

www.researchonline.mq.edu.au

Prepared for EDUCAUSE Australasia 2009

Macquarie University

Macquarie University is located in the north-western suburbs of Sydney, NSW, Australia, and was founded in 1964. Macquarie University's goal is to be among the top eight research universities in Australia and one of the top 200 in the world by its 50th birthday in 2014, . Further information is available at <http://www.mq.edu.au/university/index.html>

In 2008, the University had 33,000 students (20,692 full-time and 12,090 part-time) and 2,221 staff (1,098 Academic and 1,123 Professional staff).

Macquarie University ResearchOnline

Macquarie University ResearchOnline is Macquarie University's open access digital collection. It is designed to promote globally, preserve locally and provide open access to the research and scholarly output of Macquarie University's staff, students and affiliates.

Macquarie University ResearchOnline went live in December 2007 with 148 records. As of 21st April, 2009, there were 4,443 records loaded, of which 594 had full-text available. The Top 10 resource types are:

- journal article (2709)
- book chapter (588)
- conference paper (397)
- book (240)
- Brunner digitised document (173)
- newspaper article (80)
- review (66)
- thesis, PhD (55)
- conference paper abstract (45)
- still image (22)

We are loading records for all Macquarie research publications, even if no full-text is available, as we believed that it was important for *Macquarie University ResearchOnline* to be a full record of all the University's outputs. For more information about the Collection Development Guidelines, see <http://www.researchonline.mq.edu.au/vital/access/manager/About>

Repository management structure

Macquarie University ResearchOnline is currently managed by the Library's Digital Repository Project, which is part of the Library's Project Office. The repository is installed on a server located in the Library's Library IT Department.

Partnerships, collaborations

Macquarie University ResearchOnline has been a result of collaboration with the University's Research Office as the main source of metadata for the Macquarie University ResearchOnline has been the HERDC data from the University's Research Management system.

Macquarie University Library was a member of the RUBRIC and ARROW Projects whilst they were in existence.

Name of the repository – any tips for others on choosing a name

The repository name is "*Macquarie University ResearchOnline*" or "*ResearchOnline*" for short.

When we were looking at names we had many things it shouldn't be: no clever acronyms, nothing starting with an 'e' or 'publications'. We then looked at what we wanted the name to say and we agreed we were advertising the university, and it was to showcase our research online. We discussed whether we should have the university name as part of the name, at the front or end of the name, or to leave it out. However we all agreed that as it was to advertise Macquarie University's research online, we had our name.

Regulatory environment within your institution

Macquarie University passed an open access mandate on 27th August 2008 which requires:

- All refereed, revised, final draft research manuscripts will be deposited in the Macquarie University Digital Repository after their acceptance for publication Books or chapters in books may be self-archived at the author's discretion.
- These manuscripts will be made open access, available to anyone on the Internet, except where this is restricted by publisher policy.

The full details of this policy is available at http://www.mq.edu.au/policy/docs/open_access/policy.html

The Library is working with the Research Office and the Faculties on the implementation of this mandate during 2009. Some of the University's Associate Deans for research are starting to include open access deposit as part of an academic staff member's annual performance targets.

Models

Institutional model based on the Arrow HERDC report

- <http://www.arrow.edu.au/docs/files/arrow-herdc-workinggroupreport.pdf>

Currently Macquarie University most closely matches Model 2 in the ARROW HERDC report. However not all required data for the Macquarie University ResearchOnline is captured by the HERDC process, e.g. DOI, abstract. This data, if it is not in the metadata is sourced by Repository staff.

Institutional embedding and innovative Practices

Currently *Macquarie University ResearchOnline* is run as a project within the Library. During 2010 planning will occur on incorporating the Macquarie University ResearchOnline into the Library's

standard Organisational structure. It is anticipated that any organisational change will occur from 2011.

The Library is also anticipating working with the University's Research Office on loading data from the Macquarie University ResearchOnline back into the Research Management system. This will be initially trialled for the 2009 ERA process. Whilst dealing with the ERA requirements is a substantial amount of work, it is our chance to show the role of *Macquarie University ResearchOnline* in promoting research at Macquarie.

Outreach, Marketing and Faculty Engagement

Strategies and Services

Whilst we conducted some outreach and marketing within the University in 2008, this will be a major activity for 2009. The priority will be to sell *Macquarie University ResearchOnline*, the open access mandate, so we can get researcher involvement in providing us with open access digital versions.

The Repository project staff are currently working with the Library's Information Access (liaison) staff to assess how this outreach and faculty engagement can be strengthened by further involving these staff. They are the ones with the excellent connections to the faculty and we need to take further advantage of these connections.

Frequently asked questions

We have developed a series of Frequently Asked Questions relating to the Macquarie University ResearchOnline and open access in general. These are available at:

<http://www.researchonline.mq.edu.au/vital/access/manager/Faqmq>

<http://www.researchonline.mq.edu.au/vital/access/manager/Faqoamq>

Copyright, licencing, (include use of Creative Commons) and legal aspects

The Repository staff undertake all copyright checking for the authors. The Repository Project, in conjunction with the University's Copyright Coordinator and the University Solicitor, developed a Depositor's licence which we use for the deposit of open access versions. However the requirement for this licence is under revision as a result of the Open access mandate.

We use the SHERPA and OAKList for sourcing copyright information and rely on them greatly in our processes. Our ability to contribute so far to the OAKList has been limited by our need to concentrate on the ERA publications.

It would also be very useful if there could be suitable copyright training sessions developed for repository staff on issues associated with repositories. There are some great copyright sessions for libraries relating to CAL licences, usage by educational institutions etc, but nothing for repository staff. I know certain processes will be institution-specific but I am sure there would be some general training that could be developed.

Barriers to moving forward

Whilst ERA is of great advantage to the Repository project as it is helping to develop the awareness amongst our researchers of the existence of the repository, we are also finding that the requirement

to process the older ERA material is limiting our ability to concentrate on sourcing open access versions for all the current publications.

We are finding there is a great lack of knowledge with a number of publishers where they don't understand what repositories are, and why we want to archive a copy of the publications.

Relationships with Research Management Systems and other systems

Interoperability, data exchanges, and ongoing relationships

We are working with our Research Office on the best way of extracting data from our Research Management system. Currently we are trialling the use of data from an automated reporting system which allows the extract of data from Research Master for conversion into the format we require for our work.

Data duplication and/or matching

We avoid data duplication where possible. Our aim is to have the data created once and we re-use it with as little editing as possible. For our research publications, we use the data entered into the University's Research Management system. We consider the Research Management system as the self-submission system for *Macquarie University ResearchOnline*.

For our digital theses and other similar data, we use the data from our Library management system.

Single signon and authentication issues

Currently *Macquarie University ResearchOnline* has no authentication options available as it was designed to be an open access repository. However work is about to commence on implementing the authentication required for the 2009 ERA Trial.

The requirement for an authentication is a major system issue that we have to address in 2009.

Output to personal web pages, CV creation, Bibliographies and Social Networking tools

We have trialled the exporting of data to personal web pages and this does work. However this feature is hampered by the many variations of author names. We would hope that the ARROW NicNames project will resolve this issue and we can take advantage of their recommendations. Developing other tools such as these are areas for future development.

Research Assessment and Reporting

Repository support for formal research assessment exercises

Macquarie University ResearchOnline will be used as part of the Macquarie University's ERA submission for 2009 and onwards. Over the next year we will be looking at its role in supporting other formal research assessment exercises.

Technical Environment and Information Technology strategies - Platforms/Architectures

Software Platform(s)

Red Hat Linux on a VM server

Database and Operating Systems

Fedora Database with VTLs VITAL software

Customisations – how heavily is your system customised?

We have done a certain amount of customisation but we also believe that the majority of the usage of *Macquarie University ResearchOnline* comes through sources such as Google, so it is more important to have good quality metadata and full-text versions rather than an extremely customised site.

However we have added some local indexes which we think will be useful for our local usage.

Metadata schemas

We use MARCXML with Dublin Core as our display metadata.

Harvesting

We are OAI-PMH compliant and have PROAI implemented. We are being successfully harvested by the ARROW Discovery Service, Google and other open access sources.

Sustainability

Organisational Structure - Staffing and resources

Macquarie University ResearchOnline is managed as a project and reports to the Deputy University Librarian.

The current staffing structure is:

Digital Repository Project Manager

Digital Repository Project Officer

Digital Repository Project Librarian (casual)

Digital Repository Project Support Officers x 4 (3 until Dec 2009, 1 until June 2009)

All positions are secondments or contracts until December 2009.

Funding and external grants

The project is funded through the Australian Scheme for Higher Education Repositories (ASHER), together with some funding given by the Vice Chancellor for the implementation of open access. The project is currently funded until December 2009.

Management structure – reporting lines

The Repository Project and *Macquarie University ResearchOnline* is managed as a project and reports to the Deputy University Librarian.

Rate the quality of your metadata

I would hope the metadata quality is high as Repository staff have verified all records before loading into *Macquarie University ResearchOnline*. We have enhanced the records with abstracts and keywords where available.

There will be some normal inconsistencies of data where we have changed our standards along the way but hopefully the metadata is reasonably accurate.

How much Open Access full text does your repository contain? – is it a question of resources or copyright?

Currently the Macquarie University ResearchOnline contains 14% full-text. However once you exclude formats, i.e. books where an open access digital version is unlikely to be available, the % increases to 17%.

The low full-text rate is due to the need to concentrate on the ERA records for 2002 onwards. For publications from 2002 – 2006, the decision was made not to request open access author manuscript versions, because of the low rate of supply from the authors for the older material. For 2007 publications onwards, authors are being asked for manuscript versions where copyright conditions permit their loading.

Once the 2009 ERA work is complete, work will revert to processing publications from 2007 onwards and sourcing open access versions where possible from the author.

Lessons learned and Future Plans

What is next? 1 year? 2 years? Longterm

In 2009, we need to ensure we meet the ERA2009 Trial requirements. Our priority is also to start promoting the open access mandate and working with our researchers to source open access versions. This is our make or break year where we prove to the Library and the University what we are doing is worthwhile and so worth supporting and so ensure we guarantee our ongoing funding.

In 2010, we need to meet the full ERA requirements. However as we only have funding guaranteed for 2009 we will be undertaking as much of the ERA work this year as possible. I am also hoping we can extend *Macquarie University ResearchOnline* so it is recognised as an excellent open access resource of Macquarie University's research. I would also be hoping that our open access mandate is widely known amongst our researchers and accepted as a useful means of promoting their research.

In 2011, *Macquarie University ResearchOnline* should be fully integrated into the Library's Organisational Structure and have a guaranteed ongoing budget and staffing resources.

What has been easy to solve?

The support from Library Executive has been excellent with both the University Librarian and Deputy University Librarian aware of the challenges and issues associated with the project and acting as champions for the project. The achievement of an open access mandate was also achieved through the support of the Vice Chancellor. Having such senior support and promotion of what we are doing, will be of a massive advantage as we start to promote *Macquarie University ResearchOnline* this year.

What has been difficult to solve?

Designing our processes to fit in with the other University research publications processes and what is the best way to source our data that involves the least amount of double handling as possible.

Developing our metadata standards and requirements for the various uses whilst the data keeps changing.

What advice would you give others?

You are not alone. Most repository managers are coping with changing priorities, standards, systems every day and often you feel like you are making it up as you go along. Whilst we are using different software, the big issues such as copyright, metadata and standards are software independent. So find a group of repository managers to talk to, ask questions of, and share information with. In that way we can all advance.

Fiona Burton
Digital Repository Project Manager
Project Office
Library
MACQUARIE UNIVERSITY NSW 2109

Phone: +61 (0)2 9850 7556

Fax: +61 (0)2 9850 7568

fiona.burton@mq.edu.au

www.library.mq.edu.au

www.researchonline.mq.edu.au

Murdoch University's Research Repository: a case study in persistence.

Joanne Richards, Rob Rochester, Ploy Tangtulyangkul, Del Shiers, and Sue Hele

Prepared 30th March 2009

Abstract/Summary

The Murdoch University Research Repository has been a long time in the planning. A number of issues converged to make the development process seem endless to those that were there in the beginning when the Library joined the RUBRIC project in 2006 and for those of us who are more recent additions to the repository team. Whilst the delay has been frustrating for all concerned we are now starting to see the benefits of taking our time. Problems and considerations that have been highlighted by the repository community over the last three years have been taken on board by us. Despite the technical, financial and staffing delays and difficulties that we have had over that period we feel that the end result is shaping up to be something that we can be proud of. This case study outlines some of the considerations and problems that we have had to face, as well as some of the successes that we have had.

Institutional Overview

Murdoch University first opened its doors to students in 1975 and in the period since then has been committed to providing excellence in both teaching and research. Currently with over 8,500 FTE students and 499 FTE academics, Murdoch has developed a flexible admissions policy that provides alternative entry paths for students from non-traditional academic backgrounds and an adaptable undergraduate degree structure which allows students to broaden their qualifications. Success in achieving the teaching and research goals can be measured by achievements such as receiving a 5-star rating for Graduate Satisfaction for 13 out of the last 14 years and being ranked in the top eight universities in Australia for research intensity.

Murdoch University's Research Repository is managed from the Library and our involvement commenced with participation in the RUBRIC project as a cross-sector partner (IRUA) via University of Newcastle Library. Involvement started in 2006 and continued until the project ended in Dec 2007. The timelines and objectives of the RUBRIC project were adopted by Murdoch University.

Since the completion of the RUBRIC project there have been delays in establishing the repository. These have mostly been as a result of software, staffing and funding issues. A lack of IT expertise in the Library, which is a consequence of the university policy to have centrally located IT Services, hindered the resolution of problems associated with setting up the VITAL software. The delay in having ASHER funding confirmed meant that it was several months before staff could be employed to work on the setting up of the repository.

Models

A model for the integration of HERDC and institutional repository processes has not been officially established as yet, but we are working towards Model 3: Shared Input.

The University is in the process of implementing a new research management system (IRMA) that will enable researchers to submit research publications for the HERDC on a continuous year-round basis (as opposed to the current 'end of collection period' basis).

When the continuous HERDC collection process is fully established, both the Division of Research and Development and the Library have agreed that it would be desirable to establish a shared input pathway via IRMA to reduce time demands on the authors and to increase the opportunity for the repository to capture research outputs in a timelier manner. Work on a shared input tool has not commenced yet.

Outreach, Marketing and Faculty Engagement

The Murdoch Research Repository is still in the development phase and at present only contains research theses, which are successfully being harvested by ARROW Discovery Service. We have been working towards adding other research items and have recently resolved software and copyright management issues that have prevented us from adding other items. This period of hiatus has had the benefit of allowing us to give considerable thought to the structure and appearance of the repository, and how we are going to address the ongoing problems and issues that seem widespread within the repository community.

At present the Murdoch Research Repository is being funded by ASHER. Funding beyond 2009 has not yet been resolved due to delays in re-establishing a Repository Steering Committee following the resignation of key staff members. As a result, it has been decided to keep the aims of the repository small until we have a clearer idea of what the future funding situation will look like.

To this end, we are planning to initially limit the scope of the repository to only theses and any HERDC eligible publications for which a full-text file (that can be made publically available) has been obtained. We are not intending to create metadata records for publications unless a user can download a full-text file from the repository.

To make the process of contributing content as easy for researchers as possible, we are intending to adopt a highly mediated workflow in which repository staff will undertake most of the tasks involved in 'depositing' a work to the repository.

We will ask authors to forward a copy of their 'post-print' file to us, either:

- via an email sent to a central repository email address, or
- via a cut-down online deposit form that in most cases will only require an author to login via the university authentication system (thus automatically embedding their name and ID number into the form) and then upload their publication using an attach file command. There will be a box to provide more information, which we envisage will only be required if sufficient identifying information is not on the item that they have deposited.

The process of actually creating the metadata record in the repository system, of checking copyright permissions, and converting files to PDF (if applicable) will be solely undertaken by repository staff.

While management issues are being resolved we felt it would be preferable to have a 'soft' launch of the repository. During the periods of delay we have been working through the university's HERDC data from the last three years, and in the process, have come across a significant number of items where the publisher allows their formatted PDF to be deposited into the repository (sometimes

with embargoes). We have access to a vast majority of these publisher PDF files via our subscription databases.

We have decided to use these items to introduce the repository to our research community. We will email staff and let them know that we have permission to make one of their publications available in the repository, and all they need to do is send back to us a deposit licence form which will be attached to the email. Once repository staff have created the metadata record and uploaded the publisher version, we will contact the researcher again and both inform them that the article is available, and inquire if they have kept the 'post-print' versions for any of their other publications. We hope that by approaching them in this way we will be able to establish relationships that will encourage further use of the repository. We feel that these relationships will be enhanced by our plan to do the work involved in depositing and establishing copyright permissions for the items.

Permission from the authors will be by way of their signature on the Research Repository Deposit Licence Agreement. One of the more recent delays in setting up the repository has been caused by negotiating the licence requirements that our researchers must sign. The Office of Legal and Governance have in the past week agreed to a licence for our depositors that only needs to be signed once by our research staff. We feel our success in having this agreed to will prove to be a very positive step towards encouraging our researchers to use the repository.

We do not have an FAQ page and are not planning one in the near future. We have information about the repository on our About, Copyright, and Contribute pages. As we are doing the work of depositing we do not feel extensive FAQ's are necessary.

The unknown level of institutional support for the repository is presently the largest barrier to moving forward. This uncertainty has meant that our decisions so far have reflected the possibility that there may not be enough support in the future to do all that we would wish with our repository. We hope this is not the situation, but felt that it would be easier to increase our aims and goals should significant funding be received. If the desired level of support from the university is not forthcoming we will have a small, but well planned repository that showcases the work of those researchers who are enthusiastic about it as a means of preserving and making openly accessible their work.

Technical Environment and Information Technology strategies - Platforms/Architectures

The Murdoch University Repository's technical information is as follows:

- The Software Platform used is VTLs VITAL 3.1.1 running on Tomcat
- Oracle database and Redhat Operating System.
- Unique setups that have been put in place include:
 - Uses Apache to bypass tomcat port via mod_proxy
 - Fabulous – for editing records
 - VALET and Squire – for depositing forms
 - OAIprovider – for oai harvesting
- Information Technology Services (ITS) does a regular backup of the systems.
- Customisations - Medium to Heavy – especially templates, add-on applications and datastreams
- Metadata Schemas - MARC and DC
- Open URL linking – not required as yet as the repository only has full text

- Harvesting – we use OAIprovider replacing standard fedora OAI.

Sustainability

The repository project currently has 2.8 FTE staff which includes a 0.8 IT Officer. At present it is fully funded by ASHER. The full time Repository Coordinator reports to the Manager of Collection Services.

A considerable amount of time has been spent on the repository metadata and we are happy with our progress in this area. One innovation that we have been able to set up has been the creation of MARC subtags in the Browse by Resource Type index. This has enabled us to separate the different types of theses. This is important as we are planning to include honours and other types of student work in our repository. The subtags allow us to harvest the ADT items separately from the other types of theses.

At this time we have 100% Open Access full text resources in our repository. We are planning to continue with this in the early stages, but may reassess going forward. The main reason why we have chosen to do it this way is because the future of our repository has not been formalised as yet. With the level of support that we will receive from the university being unknown we wanted to start out with a resource that we felt we would be able to maintain into the future. Should the university decide to provide substantial support the repository's collection policy will be reconsidered.

Lessons learned and Future Plans

Over the next year we plan to put considerable effort into liaising with the university's researchers. While we await our official launch and marketing 'blitz' (which should take place in the next few months) we will focus on authors who we identify as having relevant items. We will also be liaising with the library's Liaison Librarians to use their expertise and knowledge about the Murdoch authors who they believe may be interested in depositing items into the repository.

We also plan to assess the VITAL software to determine whether it is fulfilling our expectations. To this point in time we have not been happy with the technical support that we have received from VTLS, but that may change with the recent opening of an Australian office. As part of this appraisal process we intend to evaluate other software that is available, so that we can better understand what our options are going forward.

Over the next two to three years we plan to work towards embedding the repository workflows so that they become integrated with other library procedures. We would also like to have the depositing of items accepted as a normal process in the university's research community's workflows. It is also our aim to work towards having the university accept a mandate for deposit into the repository. We are hopeful that this will be well received by the Academic Council as the university has a previous record of being receptive to mandates having been an early adopter of a mandate for digital deposits of PhD theses in 2003.

Unfortunately, like most repository teams, we have not found setting up a repository easy. The most difficult problems for us to resolve has been IT support, or lack of it. The central IT Services have struggled to find staff to adequately deal with the problems we have encountered while trying to

configure the VITAL software. We have also found it difficult to locate staff with expertise in working with repositories, in particular VITAL repositories.

Advice for others? With our repository still being untested in the wider world we don't feel qualified to offer too much advice, however we do feel that we have been able to gain from the delays that have been imposed upon us. Hopefully, with the benefit of being able to watch the trials and tribulations of others, we have been able to resolve most issues before our repository is widely used. Only time will tell, but we are becoming increasingly pleased with the work that we have done and are now looking forward with anticipation to the next phase, which will see us introducing the repository to the university community and populating it with their substantial research efforts.

Swinburne University of Technology
Research repository case study

Swinburne Research Bank
<http://researchbank.swinburne.edu.au>

For presentation at the Research Repository Managers Symposium at EDUCAUSE Australasia
2009, Perth, Western Australia, 03-06 May 2009

Prepared by:
Rebecca Parker
Content Management Librarian
Swinburne University of Technology
Hawthorn, Victoria, Australia
rparker@swin.edu.au

April 2009

Abstract

At Swinburne University of Technology, our institutional repository is called Swinburne Research Bank (<http://researchbank.swinburne.edu.au>). We use the VITAL interface to Fedora, and we have just over 11,000 records, 16 percent of which are full text. Our aim for Swinburne Research Bank is to create a complete picture of research published at Swinburne; by listening to the advice of our academics and broadening our collection policy to include unconventional resource types such as seminars, newspaper articles and reports, we hope to catch more of the informal research dissemination activities that happen around the University. We run a mediated deposit service and so far have successfully populated our repository without a mandate.

For a more detailed report on Swinburne Research Bank content management strategies, see:

Parker, R., & Wolff, H. (2009). Building Swinburne Research Bank: an engaged, user-centred approach to content recruitment. *Paper presented at the 14th ALIA Information Online Exhibition and Conference, Sydney, New South Wales, Australia, 20-22 January 2009*. Retrieved 24 April 2009, from <http://www.information-online.com.au/>.

About Swinburne University of Technology

Swinburne is a medium-sized, dual-sector educational institution incorporating both higher education and vocational education (TAFE) programs. There are 6 campuses in Victoria and one in Malaysia, with the main campus situated at Hawthorn in the inner suburbs of Melbourne. Swinburne celebrated its centenary as a technical college in 2008 and has a consistently high rating in teaching and learning, but there is a strong focus for the future on developing the research culture of the University. Swinburne is a member of 4 ARC Centres of Excellence in areas including quantum optics, microphotonics, x-ray science and creative industries.

About Swinburne's repositories

The research repository at Swinburne University of Technology is Swinburne Research Bank (<http://researchbank.swinburne.edu.au>). It was developed with the assistance of the Australian Research Repositories Online to the World (ARROW) Project, and uses the VITAL interface to Fedora developed by VTLS Inc. in conjunction with ARROW.

The repository is managed by the Online Services and Strategies Unit at the Library, which means that repository staff are co-located with other online projects staff such as the Library website coordinators and the Library user experience consultant.

At Swinburne we have made a conscious decision to run multiple institutional repositories where there is a need to keep different content types separate to ensure their cohesiveness and clarity. Consequently the Library maintains an image database separately from the research repository. Swinburne Image Bank (<http://images.swinburne.edu.au>) contains mostly archival multimedia and teaching images, and uses open-source DSpace software from MIT.

Systems and metadata environment

Repository systems

- * Fedora
- * VITAL 3.3.4 incorporating usability customisations recommended by Dana McKay [3]
- * Oracle database
- * Active Open URL resolver (SFX)

Repository harvesting

- * OAI-PMH compliant
- * Harvested by OAIster (<http://www.oaister.org/>), the ARROW Discovery Service (<http://search.arrow.edu.au/>) and Scientific Commons (<http://www.scientificcommons.org/>)
- * Indexed by Google and other web search engines

Repository metadata

- * MARCXML as underlying metadata schema
- * Dublin Core for item display metadata

About Swinburne Research Bank

Timeline and aims

Swinburne Research Bank first went live in 2006 and was publicly launched to the Swinburne research community in 2008. Our aim for Swinburne Research Bank is to create a complete picture of research published at Swinburne. We were concerned that in concentrating our attention solely on HERDC-eligible research outputs, such as journal articles, books and conference papers, we might exclude a large part of both the formal and informal scholarly communication and dissemination of our academics. HERDC rules not only restrict content to set formats, but also prohibit the acceptance of research results published in a variety of formats, for example a positioning paper, conference paper and then journal article that might all represent the results of a single experiment at various stages. For many disciplines, it is important to be able to see this progression from early ideas to final conclusions to fully understand the impact of the research. By broadening our collection policy to include unconventional resource types such as seminars, newspaper articles and reports, we hope to catch more of this informal and early dissemination activity within Swinburne.

For a full list of publication types currently available from Swinburne Research Bank, see our Browse by Publication Type menu:

<http://researchbank.swinburne.edu.au/vital/access/manager/Browse/Type>

Name

Our broader vision of Swinburne's research repository as the home for all Swinburne research also influenced our choice of name for the service. At Swinburne Library we have a usability and user experience consultant who is able to advise on the appropriate and clear use of terminology for our services. The brand 'Swinburne Research Bank' is deceptively simple; in fact, a number of considerations went into choosing this name. While it is too long to be catchy, it avoids any of the terms such as 'repository', 'institutional repository', 'digital repository' and 'digital library' that the literature indicates our users might find confusing. It is also software-independent; many universities have used 'EPrints' or 'DSpace' in the title of their repositories, leaving little room for major developments in repository technology. While Swinburne's repository is (like most institutional repositories) in fact maintained by the Library, we deliberately chose not to add the Library brand to the repository, preferring instead to associate it more broadly with the University as a whole. We also elected not to use 'ARROW' in the title, as the project had a finite end date—one that we hoped Swinburne's institutional repository would significantly outlast.

We believe the semantic link between the name *Swinburne Research Bank* and research at Swinburne is vital to the success of the repository. University of Rochester academics believe the term 'institutional repository' 'implies the system is designed to support and achieve the needs and goals of the institution, not necessarily those of the individual ... [and] contributions of materials into the repository serve to highlight the achievements of the institution, rather than those of individual researchers and authors' [1]. This is a disconnect we wished to avoid and would recommend that other repository managers bear this in mind when choosing a name for their repositories. The name is more than just a title for the system—it's a brand for the whole service.

Collaboration and partnerships

Swinburne was an original partner of the ARROW Project, and consequently repository staff have built and maintained valuable collaborative partnerships with other ARROW repository managers.

Repository staff also have strong ties with other corporate areas of the University, especially the research office, Swinburne Research. These partnerships have allowed us to make the repository responsive to the needs of both academics and the University as a whole. As a consequence, multiple services and projects reuse repository publications data---a mark of the trust that the University has in the value of Swinburne Research Bank.

Models for research management system integration

In the early days of populating our repository, we received all information about HERDC-eligible publications from our research office at the end of each year's return. This model closely followed Model 2 as defined by the ARROW HERDC Working Group in their report on integration workflows [2]. However, for the 2009 collection of 2008 publications, we are trialling a system that more closely follows Model 1. Swinburne Research Bank staff create a bibliographic record for all HERDC-eligible publications, then export that information to Swinburne's research management system for the next step of the process. Verification material is centrally stored and accessible to both repository and research office staff. So far this process has been very successful and saved much of the double-handling of publications information that occurs when the two systems work completely independently. We will review the process at the end of the collection period.

Reuse of Swinburne Research Bank data

In addition to our role in the HERDC process, Swinburne Research Bank data is reused in a variety of settings around the University. We are able to provide a personalised URL for academic staff to link to their publications in Swinburne Research Bank, and we are investigating working with faculties to provide them with feeds to dynamically update their publication pages.

Swinburne Research Bank staff also collaborated with the Swinburne Marketing department and research office to automatically generate profile pages for all Swinburne researchers. The profile pages began as a project funded by ARROW to investigate how data from existing university systems (for example, identity management systems, research management systems and research repositories) might be reused to create dynamically-updated profile pages for researchers. It is likely that the profile pages will play a significant role in the way Swinburne academics are represented on the Web.

Outreach, marketing and faculty engagement

Swinburne's approach to content recruitment

At Swinburne, we know we do things differently. One of the most significant ways we deviate from standard practice with our repository is by rejecting self-deposit, at least as the primary means of content recruitment for our repository. This means that our service is highly personalised; we very closely monitor any research that is published with a Swinburne affiliation through a wide range of media including library databases, Libraries Australia, internal university publications and Google Alerts (<http://www.google.com/alerts>). We also receive regular reports of new academic staff from our Human Resources department.

From this information, we are able to create bibliographic records for publications before we even begin to contact researchers about the repository. This means that on their first glance at Swinburne Research Bank, most Swinburne researchers should see at least some of their own publications. We believe that this individual attention gives stronger incentives for everyone to contribute to their repository, from early career researchers to faculty deans.

Know thy discipline

One of the many very useful facts we have learned over the years is the impact of discipline on the repository and the way we should promote it to academics. While the open access repository literature always stresses the effect of repositories on multiplying citation counts, some disciplines predominantly publish outside the journal culture and consequently don't measure the impact of their research by numbers. For example, while it may be helpful to push the citation increase angle with scientists, many social scientists and humanists are more interested in making their work more widely available through Google than the effect this increased exposure might have on their citation metrics.

Listen to the academics, not the literature

One of the reasons that academics can be reluctant to contribute to an institutional repository is the perception that they contain sub-standard drafts that misrepresent their authors, and not an archived copy of quality research. This may partly be confusion over the concept of the author's final draft, or accepted manuscript, which can also be known as a postprint. Authors often mistake the term 'postprint' for 'preprint'. And little wonder.

Many of our academics express a preference for the published version (or version of record) of their paper to appear in Swinburne Research Bank, and would prefer not to use any other version. This is a wish that we respect. We would rather have no full text access to a paper than a draft that makes our academics uncomfortable.

Statistics

Due to delays with software development, Swinburne Research Bank does not currently display usage statistics publicly. This is a pity, as it's one of the questions we are asked most when giving presentations to academics, especially in disciplines where citation metrics are not a useful or effective measure of research impact.

We do keep internal statistics using Google Analytics (<http://www.google.com/analytics>), and these provide us with very interesting data such as most frequently accessed papers, most common search terms and the geographic diversity of our users. The most surprising discoveries so far are the continued popularity of our thesis collection, and the fact that we're visited an average of 150 times a day by users from 177 countries.

Sustainability of Swinburne Research Bank

Staffing and resources

Currently, as a result of first ARROW and now the Australian Scheme for Higher Education Repositories (ASHER) funding, we have a 0.5 FTE repository manager and 3.5 FTE content librarians working on Swinburne Research Bank. However, their time is split between the repository and other online projects managed by Swinburne Library including our image database, our preparations for ERA and HERDC submissions, and our open access publishing services.

Metadata quality

Repositories like ours that don't rely on self-deposit as the primary method of content recruitment are known in the literature as 'mediated' repositories. However, this term implies that repository staff merely intervene in the repository deposit process—in our case we are responsible for *all* metadata creation activities for the repository.

One of the benefits of having repository staff create our descriptive metadata with the support of traditional reference tools such as bibliographic databases is that we can assure consistent and high quality metadata. The University of Waikato has devised a metadata analysis tool [4] to review the completeness of repository metadata. At last check, our metadata completeness rated above 90 percent, despite us having well over 10,000 records created over a number of years from a variety of sources. While high quality metadata creation by repository staff might be seen as taking time away from other repository tasks, the fact that other areas of the University rely on the quality of our metadata for their public activities and our academics rarely find fault with our records indicates that the investment is worthwhile.

Open access content

At the moment, our repository consists of over 11,000 records, about 16 percent of which are full text. Where we haven't yet been able to source the full text of a publication, in most cases we provide a secondary path to full text through a DOI link or URL. However, we are focussed on building our hosted full text content in 2009 and 2010.

We are not alone in our modest full text access, and we believe this relates to both resources and copyright, not either in isolation. While services such as SHERPA/RoMEO [5] and the OAKList [6] provide invaluable information on publisher stances on open access, their focus has always been on journal article copyright and not on conference papers or book chapters. In the first four months of 2009 alone, Swinburne Research Bank staff have contacted over 100 publishers individually to ask for permission to archive full text, either because the information wasn't available through the databases or the record didn't specify the option to store the published version (our preferred version of the full text to archive in Swinburne Research Bank).

It is not uncommon for staff to devote hours or even days to the sourcing of full text permission for conference papers. In many cases a conference website will disappear overnight, and this time is spent trying to hunt down anyone vaguely connected to the event, including emailing proceedings editors who usually are happy to endorse open access to the publication once we finally track them down. As a possible recourse to this investment of time and labour, we are developing a risk management strategy for open access to conference papers, taking into account whether the paper has been published in a proceedings, whether the website still exists and whether the original publisher can be traced. We believe this will be a useful tool for other repository managers in the future.

Future plans for Swinburne Research Bank

- * Increasing full text to 25 percent
- * Further streamlining the integration of the HERDC process with the repository
- * Launching additional services such as profile pages to the Swinburne community
- * Providing advice to researchers on research assessment and outlets for publication

Lessons for the repository community

Many repository managers claim that the hardest task in creating a successful repository is attracting the interest and support of academics. This is the primary reason that they support the need for a university-wide mandate for self-deposit. At Swinburne, this certainly hasn't been a problem for us. Many academics across disciplines quickly see the benefits of open access and believe its time has come. We have met very few barriers in the promotion of our repository that could not be overcome with further dialogue about the values and impact of open access.

However, time constraints have taught us that it pays to work closely with researchers who are enthusiastic about the benefits of open access, as their positive experience often convinces their (sometimes more reluctant) colleagues to contribute their research too. And for those who still aren't convinced about the value of a research repository after all of this attention, we believe a mandate from University administration is unlikely to make much difference.

It is no secret that institutional repository software is at present far from perfect, and there is a danger in repository managers compromising their services to match the inadequacies of their software solutions. So many of our decisions and networks are driven by the software we use, and it is often difficult to open up the dialogue with repository managers using different software but facing the same universal problems. This is a gap that we hope will be filled by CAIRSS, the software-independent CAUL Australian Institutional Repository Support Service [7].

We still have a long way to go---in some ways the ingrained library standards and practice that many of us have in our backgrounds can impair our ability to be creative with our repositories. While it is important for us to acknowledge and welcome our role in research assessment exercises such as ERA, our involvement in the provision of data for the exercise is in danger of driving the future direction of Australian research repositories. We need to focus on the needs of our immediate users—that is, our own academics.

References:

1. Fried Foster, N., & Gibbons, S. (2005). Understanding faculty to improve content recruitment for institutional repositories. *D-Lib*, 11(1). Retrieved 25 March 2009, from <http://www.dlib.org/dlib/january05/foster/01foster.html>.
2. Morgan, T., Lang, A., Wood, M. et al. (2007). ARROW HERDC Working Group final report. Clayton: ARROW Project. Retrieved 24 April 2009, from <http://arrow.edu.au/docs/files/arrow-herdc-workinggroupreport.pdf>.
3. McKay, D. (2007). VITAL 3 usability analysis: expert review. Hawthorn: Swinburne University of Technology. Retrieved 24 April 2009, from <http://hdl.handle.net/1959.3/44113>.
4. Nichols, D. M., Paynter, G. W., & Chan, C-H. et al. (2009). Experiences in deploying metadata analysis tools for institutional repositories. *Cataloging and Classification Quarterly*, 47 (3/4). Retrieved 24 April 2009, from <http://dx.doi.org/10.1080/01639370902737281>.
5. University of Nottingham. (2006). SHERPA/RoMEO: publisher copyright policies and self-archiving. Retrieved 24 April 2009, from <http://www.sherpa.ac.uk/romeo/index.html>.
6. Queensland University of Technology. (2009). OAKList database. Retrieved 24 April 2009, from <http://www.oaklist.qut.edu.au/>.
7. CAUL (2009). CAIRSS: CAUL Australian Institutional Repository Support Service. Retrieved 24 April 2009, from <http://cairss.caul.edu.au/>.

The University of New South Wales Research Repository Case Study
EDUCAUSE Australasia 2009
Prepared by Maude Frances and Joanne Croucher

Institutional overview (2008)

The University of New South Wales
42,360 students
6359 staff
8 faculties, 56 schools, 89 centres, 11 institutes, 2.7M items in library

Repository overview

Repository name: UNSWorks
URL: www.unsworks.unsw.edu.au

UNSWorks is managed by Library Repository Services (LRS) in the University Library. LRS comprises three staff: manager, support officer, and technical support officer.

Developed under the ARROW project, UNSWorks was implemented and launched in December 2007, following an ARROW repository trial during 2006-7. UNSWorks content is publicly available under a Creative Commons Attribution-NonCommercial-NoDerivs Licence.

Content (April 2009)

4140 items (2570 full text)

Journal articles	1284	Working papers	182
Reports	294	Moving images	5
Books	14	PhD theses	1522
Book chapters	67	Masters theses	295
Conference papers	421	Other resources	21
Conference items	20		

Outreach, marketing and faculty engagement

Outreach Librarians are responsible for marketing UNSWorks and liaising with UNSW academic staff on matters relating to depositing and accessing material in the repository. The UNSWorks website provides access to UNSWorks policies and guidelines as well as information relating to copyright and licensing, context-specific help and FAQs.

During development of the repository, ARROW project staff liaised closely with academic staff across various faculties. Development commenced with a series of consultations with academic staff across various disciplinary fields. Three key design requirements and priorities were identified.

Developments need to be closely aligned with existing academic publication and dissemination workflows. Data entered into the repository should be able to be re-purposed, including for display on researchers' personal webpages. The third priority was to maximize academic autonomy, responsibility and sense of ownership relating to research output in the repository.

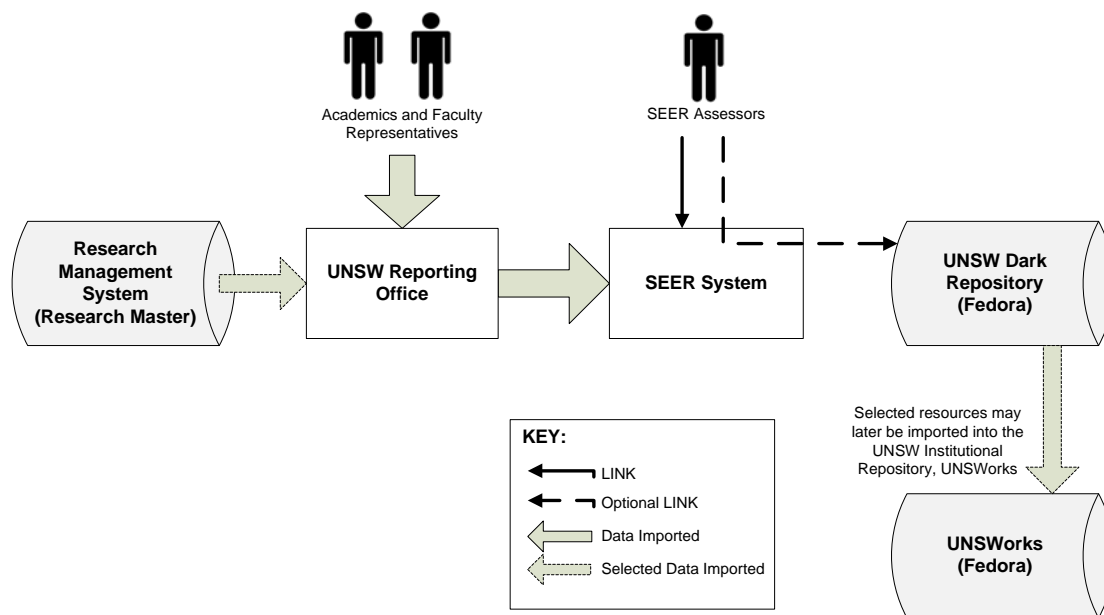
UNSWorks and HERDC collection

At present our HERDC and Institutional Repository are independent systems, with data entered, respectively, via Research Master and a web-based VALET submission form. Metadata exported from the HERDC collection to an excel spreadsheet is ingested to UNSWorks, on request from academics. A new system, due for implementation in late 2009, most closely resembles Model 3, with shared input but independent review and verification processes for the repository and publications collection.

Research assessment and reporting

A 'dark' version of the UNSW repository will be used to store research outputs for the ERA.

Figure 1: UNSW ERA System Architecture



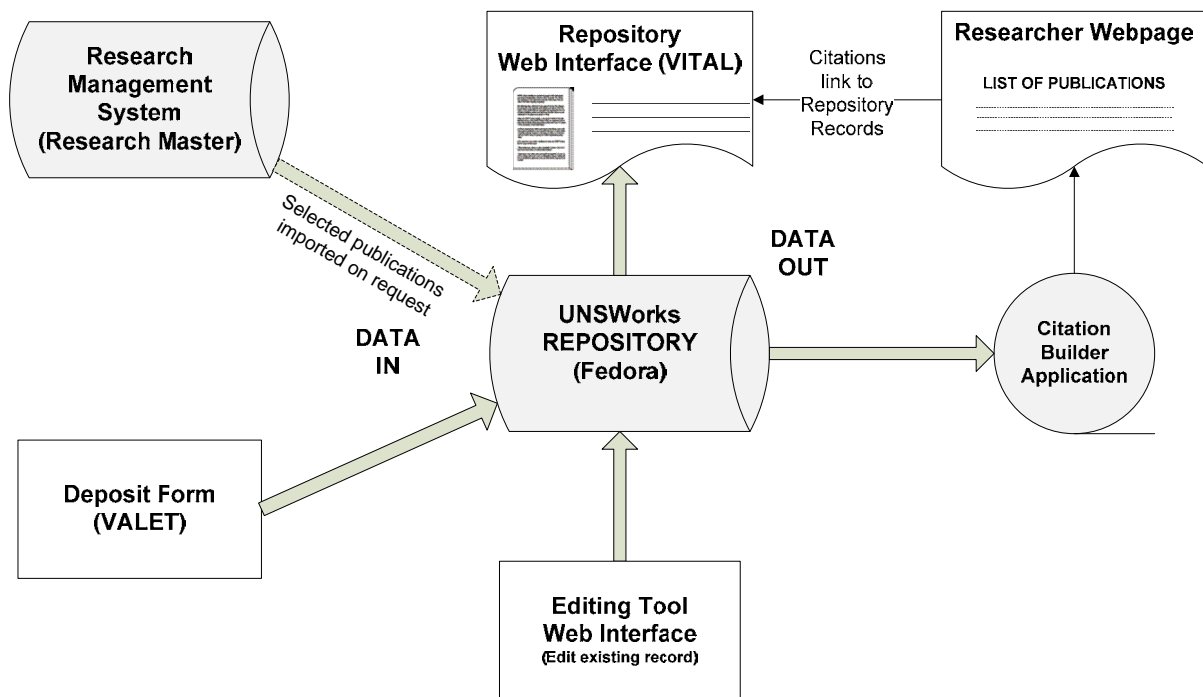
Technical environment and information technology strategies - platforms/architectures

UNSWorks is composed of a VITAL interface and a Fedora-based repository. Research outputs are represented in two descriptive metadata schemas: Dublin Core and MARCXML. Handles are used as stable URLs to individual resources in the repository. New resources are submitted using a VALET deposit form.

Due to a VITAL software problem, UNSWorks does not currently record statistics for public viewing. When the problem is fixed, in forthcoming software versions, UNSWorks will provide access to statistics on hits and downloads.

The *Editing Tool* and *Citation Builder* (outlined below) are open source applications, developed at UNSW under the ARROW project. The *Editing Tool* enables authenticated authors or faculty-based administrative staff to update repository metadata. *Citation Builder* delivers repository content, as formatted citations, directly to academic webpages.

Figure 2: Current UNSW Repository Architecture and Relationships



Other UNSW research repository projects

Citation Builder

The *Citation Builder* application enables the display of dynamic lists of publications on academic webpages. When embedded in an external webpage, *Citation Builder* uses the latest data in UNSWorks to dynamically generate formatted citations. The application can be readily implemented by web administrators, and does not require any knowledge of programming.

The initial version of *Citation Builder* software was written using PHP technology. The Java-based Version 2 is Open Source and available via Google Code:
<http://code.google.com/p/unswlibrary/downloads/list>

At the client side, three files are uploaded and a few lines inserted into the HTML. Editing this HTML enables publications to be selected for display based on specific criteria, for example, all publications by a particular author, or all PhD theses completed within a particular school of the University.

Citation Builder is highly configurable. Other Fedora repositories could configure the application to fetch the relevant descriptive metadata from Fedora (e.g. Dublin Core; MODS). While Harvard is the default citation style, the XSL can be modified to display citations in other styles.

Editing Tool

The *Editing Tool* is Java-based open source software that enables academic staff to manage their research output in UNSWorks. Authenticated authors or faculty-based administrative staff are able to update repository metadata and to add and replace digital objects. While reducing UNSWorks administrative staff workload, the application is primarily designed to maximize academic autonomy, responsibility and sense of ownership relating to research output in the repository. The Editing Tool will be tested and implemented at UNSW in May 2009.

Social Science Project

The Social Science Project is an e-research infrastructure project designed to support a multi-disciplinary research program comprising partners and stakeholders from health and social science disciplines, government and community-based organisations, and affected communities. A collection of digital resources relating to Australian social and policy research in HIV and hepatitis C, exported from relational databases developed at a UNSW centre, is ingested into a Fedora-based repository. The repository aggregates material from throughout the research lifecycle, including research proposals, project descriptions and research instruments from qualitative, quantitative and mixed-method studies. A comprehensive collection of metadata and digital objects from mainstream and community media publications, as well as cultural material such as film and visual artwork has been added. A schema of relationships links research material with associated policy, media, health promotion and other resources in the repository.

The repository links metadata describing research datasets with data archived in the Australian Social Science data Archives (ASSDA). The Data Documentation Initiative (DDI) metadata schema is applied, alongside DC and MODS, to research material in the repository.

The repository is designed to interoperate with institutional repositories. If there is an existing digital object in UNSWorks or the repository of another university, the social science repository will house a metadata only record, and provide a link to the object in the institutional repository.

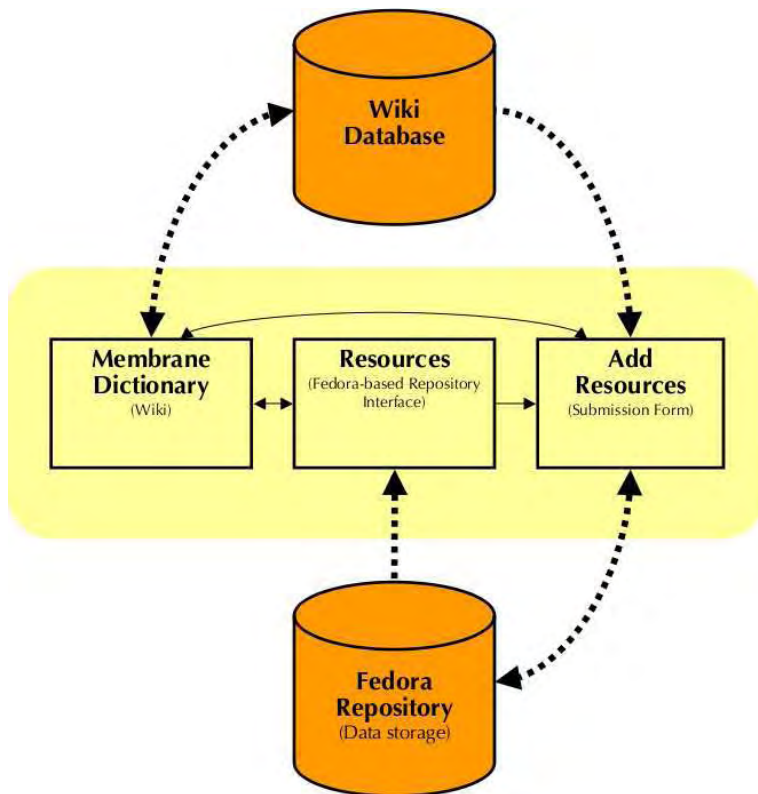
MemRE

The Membrane Research Environment (MemRE) is a component infrastructure project of the Australian national Advanced Membrane Technologies for Water Treatment Research Cluster, whose primary goal is to develop novel membrane materials in order to reduce the energy associated with desalination by 40%. An identified need of the Cluster was the development of infrastructure to support collaborative problem-based research, including knowledge transfer across multiple disciplinary groups of the geographically dispersed researchers.

MemRE was developed in 2008 through a collaboration between UNSW Cluster members at the UNESCO Centre for Membrane Science and Technology and the UNSW Library. The purpose of MemRE is to consolidate and make accessible existing research in the field, and to accelerate dissemination of new research. It currently comprises a multidisciplinary collection of literature from the last 30 years, a collection of membrane materials that have been characterised for various physical properties, and a wiki to describe the various properties and characterisation methods available for membrane materials. It integrates Fedora repository software with mediaWiki.

Functionality of the submission process for the publications repository includes a direct link to materials in the materials repository, and the submission process for the materials repository includes retrieval of data from the wiki. Access to MemRE, via a single sign-on to all three components, is currently restricted to cluster members, however, at the conclusion of the project in May 2010, content will be made freely and publicly available online. For more information see <http://handle.unsw.edu.au/1959.4/37537>

Figure 3: MemRE Repository and Wiki Architecture



Educause 2009

Research Repository Case Study : University of South Australia

Details

Institution: University of South Australia

Repository: arrow@UnISA

Website: arrow.unisa.edu.au

Email: arrow@unisa.edu.au

Case Study authors: Jenny Quilliam, Kate Sergeant, Martyn George

Date prepared: April 2009

Institutional Overview

UniSA was founded on 1 January 1991 through the amalgamation of the South Australian Institute of Technology and the Magill, Salisbury and Underdale campuses of the South Australian College of Advanced Education. Since then, it has quickly earned a reputation as a national leader in collaborative research, has been recognised nationally for innovation in teaching and has South Australia's largest intake of international students. UniSA is committed to the creation and application of knowledge. In 2008, the University secured a record of more than \$50m in total research income.

UniSA is committed to educating professionals; creating and applying knowledge; engaging our communities; maintaining cultural diversity amongst its staff and students; and providing equitable access to education for greater numbers of people. In this way, UniSA continues to build on the long-standing traditions of its antecedent institutions. Its distinctive research profile is based on bringing together multidisciplinary teams from industry, business and the professions to work on projects that are both relevant and beneficial to the community.

UniSAs emerging research priorities are sustainability, healthy lifestyle, and modelling and managing complex systems. The University strategic plans for research include:

- Building research capacity by:
 - Aligning research strategy and workforce planning processes
 - Developing a new strategic and operational framework for our research concentration policy
 - Growing research capacity in selected areas of international importance which address major national and state priorities

- Increasing recruitment of high quality research degree candidates through a variety of mechanisms
- Growing research income by:
 - Increasing the quality and quantity of institutional Australian Competitive Grant applications while also retaining a strong focus on 'end-user' sponsored research
- Increasing the number and quality of research publications and broad impact of research output.

UniSA comprises four Adelaide metropolitan campuses with regional presence at Whyalla and Mt Gambier.

Statistics Summary: in 2007, total students were 34,603, with EFSTL of 23,301 split 18,621 undergraduate EFTSL and 4,680 postgraduate EFTSL. Total staffing (2007) continuing and fixed term of 2,266 FTE, split 1,010 Academic to 1,256 Professional staff.

Repository Overview

The University repository, arrow@UniSA is managed by the University Library. Commencing in 2006, the repository was initially developed through internal competitive funding as a project sponsored by the Directors of the Library and Research and Innovation Services. The aims of the project were to provide the framework and infrastructure for a University repository, initially of research publications. The repository collects University research publications to extend the University's ability to publicise its role in undertaking primary research and provide a means to preserve and leverage its intellectual assets. The arrow@UniSA repository development was closely linked to the nationally funded ARROW project and the government-backed impetus behind this. The repository is compliant with the OAI-PMH standard thus enabling UniSA to contribute to the global institutional repository community.

arrow@UniSA was launched in 2007, however development is ongoing, particularly in the areas of content recruitment and positioning the repository to support research assessment exercises. Ongoing development is reliant primarily upon funding through ASHER to accelerate the jointly shared objective of the Government and the University to place research outputs in an accessible digital store.

During the life of the ARROW project, UniSA contributed to the ARROW community, ARMG (Australian Repository Managers Group) and MACAR (Metadata Advisory Committee for Australian Repositories) and also undertook mini-project developments funded by ARROW. UniSA intends to participate actively in the CAIRSS community.

In choosing a name for the repository, we chose to link explicitly to the ARROW project and whilst this project is now concluded, the repository name does link us nationally to the ongoing ARROW Discovery Service.

The tagline for arrow@UniSA is Discover. Deliver. Preserve. The choice of tagline is intended to reflect the key aims of the repository.

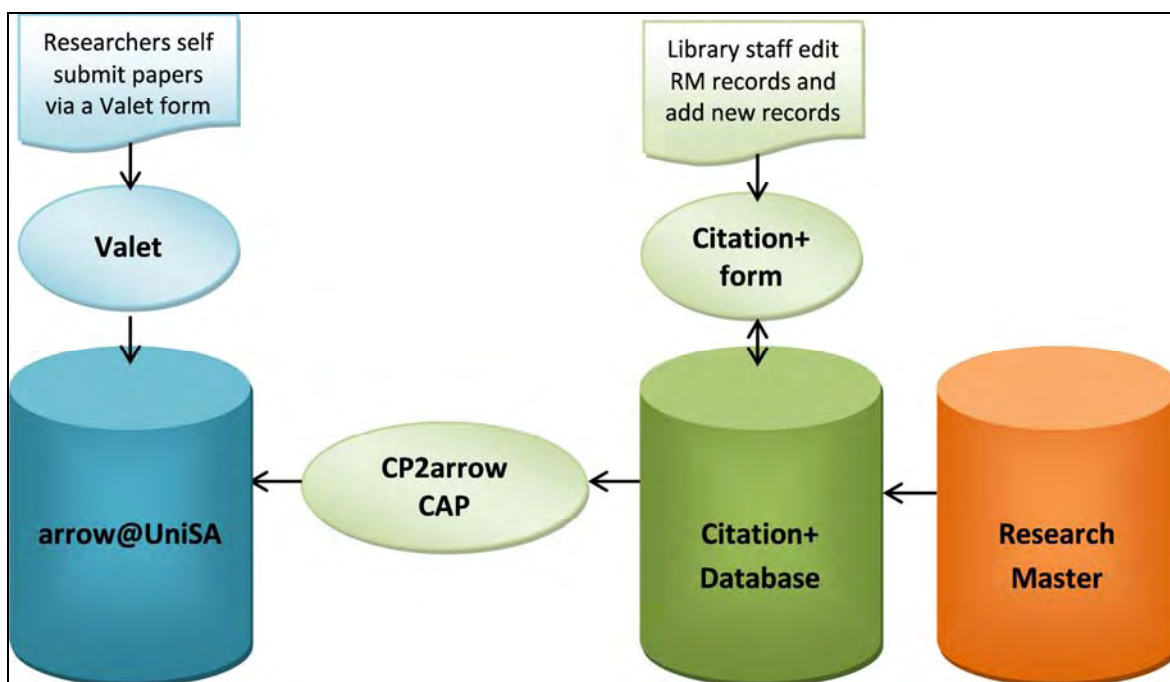
Repository statistics

As at April 2009, there are 11,831 records in arrow@UniSA – formats are text (11,281) and image (550) objects. 55% of the records include a content datastream (pdf or jpg) and 8% is open access fulltext. For records where the content datastream is inactive due to unknown or publisher copyright, a DOI is included where possible.

Content guidelines are typical of guidelines used by Australian institutional repositories, viz. works must be created, submitted or sponsored by the University of South Australia Schools and Research Centres, academic and research staff, postgraduate students, adjunct staff or visiting scholars and that the work must be scholarly, educational or related to the research or academic mission of the University.

Models

The HERDC Working Group commented that the selection of appropriate systems and workflows is always a balance between what is ideal and what is practical. Of the four models identified in the HERDC report which describe the relationship between the research management system and the institutional repository in meeting HERDC requirements, the workflow at UniSA fits most closely to Model 2 – *research management system to institutional repository*. The differences are that publications data extracted from the research management system is modified prior to ingest, and that the repository also captures and loads from other data sources. The following diagram shows the current workflows.



Outreach, Marketing and Faculty Engagement

Except for the launch which generated an initial impetus, our focus has been on collecting metadata and content in response to government research assessment exercises rather than focussing on self-submission of published outputs. However the Library's Academic Librarians do promote the repository to schools, academics commencing HDR students.

The Author services feature enables academics to create a bibliography of repository outputs for their personal webpages.

Copyright, licencing and legal aspects

UniSA uses the Creative Commons Attribution-NonCommercial-NoDerivs license (<http://creativecommons.org/licenses/by-nc-nd/3.0/>) for material deposited via self-submission. For published materials we are using both SHERPA and OAKList as well as advice from the UniSA Copyright Officer.

The most time-consuming activity in managing repository content is researching copyright status and validating content for open access. Our experience to date in negotiating copyright clearance indicates that it is easier to get clearance as an individual institution rather than clearance for all Australian repositories.

Relationships with Research Management Systems and other systems

UniSA's Research & Innovation Services uses ResearchMaster system as the University corporate information system for research related activity. The system is used to provide statutory research reporting to the Department of Education, Employment and Workplace Relations, the Australian Bureau of Statistics and monitoring research performance across corporate planning targets. The annual HERDC publications reporting is extracted from ResearchMaster.

Research Assessment and Reporting

The repository is supporting the ERA submission as a source of DOI data, supplementary source of citation data for ERA reporting plus hosting peer-review evidence. The University's ERA Submission team comprises staff from Research and Innovation Services, Information Strategy and Technology Services, Human Resources and the Library.

Since 2007, the Library has collected and harvested citation counts from Web of Science. These are currently stored in a database and will be loaded to the repository or directly linked depending upon permissions.

Technical Environment

As an ARROW implementation the repository uses Fedora / VITAL / VALET. The FABULOUS (Fedora Arrow Batch Utility with Lots of User Services), developed by UniSA IT staff, is used to assist in managing repository content. FABULOUS enables batch editing of MARCXML and subsequent DC updates and batch ingest / batch activation deactivation of repository content. A number of utilities to export / ingest data from other databases and fileshares are also in use.

Servers are managed and maintained by the University's corporate Information Strategy & Technology Services.

arrow@UniSA is harvested by the ARROW Discovery Service, ADT and GoogleScholar.

Single signon and authentication issues: Shibboleth authentication for author self-submission is operational.

Metadata

Metadata schemas used are MARCXML and DC, and the Handles system is used for identifiers. In considering the question of metadata quality, overall the quality rating is good. The focus is on deduplicating records and quality assurance of metadata. Metadata practice is based on MACAR recommendations.

The most problematic metadata issues are:

- resolving author names and identities
- inaccuracies in data extracted from other systems
- inherent problems in dealing with ephemera and grey literature

Sustainability, Organisational Structure, Staffing and Resources

The University repository has been built upon initial internal competitive IT development funding and has relied primarily upon funding through ASHER to accelerate the jointly shared objective of the Government and the University to place research outputs in an accessible digital store. Whilst development funding was secured through ARROW Mini-Projects grants, no direct funding has been received through the ARROW project to facilitate populating the repository.

The repository manager is the Library's Digital Resources Services Coordinator who reports via the Manager, Information Resources & Technology to the Director, Library Services. Technical support is provided by a Senior Information Technologist and Web Services Coordinator. For the duration of ASHER funding, additional librarians have been contracted and work in the areas of metadata quality control, content recruitment and copyright clearance.

Lessons learned and Future Plans

What is next? 1 year? 2 years? Longterm

The primary focus for 2009 is increasing the proportion of open access fulltext as required under the terms of ASHER funding and ensuring that all content required for ERA peer-review is sourced and loaded. Scopus identifiers for ERA submitted outputs will be loaded to repository records.

Mid term goals include increasing self-submission rates, work on preservation, and collection of other content types including data sets.

As repository services are now an expected ongoing digital service provided by libraries, without ongoing ASHER funding it will be important to define a sustainable service model where repository associated work is integrated into mainstream library work. Workflow optimization and ongoing inter-institutional collaboration including CAIRSS will be vital as will software to support metadata activity and manage repository content. In encouraging academic participation, as far as possible, deposit of research publications should be light touch – get it once and get it all...

What has been easy to solve?

Self submission, harvesting to ADT, harvesting to arrow discovery service. Inhouse technical & developer expertise makes it easier.

What has been difficult to solve? Managing copyright, responding to changing government research assessment exercise requirements, and technical limitations of the VITAL management system.

What advice would you give others

- Follow the QUT model of a mandatory deposit policy of research outputs
- Encourage academics (by policy) to retain author rights
- Participate in cross-institutional collaborative groups such as ARMG, CAIRSS, discussion lists, and OAKList
- Develop and maintain good working relationships with Research Office, central IT and your researchers.
- Build support with your academic library services teams
- Clean & deduplicate your metadata before ingest. Aim to load once rather than load and edit.

Research Repository Case Study

University of the Sunshine Coast

Repository Managers: Katy Watson and Daniel Viney

Repository Name: Coast Research Database

Repository URL: research.usc.edu.au

Institutional Overview

University of the Sunshine Coast was founded in 1994 with full university status granted in 1998. USC currently has 5000+ student enrolments (a growth of 11% in previous year). Of these students, over 90% are on-campus. A single campus institution, USC has 170+ teaching and research staff. Research groups, centres and facilities of USC include:

- Regional Sustainability Research Centre
- GeneCology Research Group
- Centre for Healthy Activities, Sport and Exercise
- Fraser Island Research and Education Facility
- Dilli Village Environmental Education Facility

USC research repository, 'Coast Research Database' is managed by USC Library. The Library and the Office of Research collaboratively staff and develop the repository to meet the research needs of the University. This includes integrating the repository and repository staff into research data management activities of the Office of Research (such as HERDC and ERA reporting).

The repository manager position is a joint appointment between the Library and the Office of Research - HEW 7/8 'Research Information Systems Coordinator'. Along with repository management, this role is responsible for implementing and managing the new Research Master system, and coordination of the USC ERA trial response. A HEW5 Repository Support Officer also works on the repository for 20+ hours per week.

Coast Research Database is the name of the USC research repository. This name was selected after a lengthy naming process involving the USC Marketing Division and a 3rd party Marketing Firm.

If you view the logo on the repository you can see the OA in cOAsT are highlight to represent 'open access'. This naming process fed out to the wider 'branding' of USC research as 'Coast Research' as seen

at <http://www.usc.edu.au/Research/atUSC/CoastResearch/CoastResearch.htm>.

Coast Research Database was launched in November 2007 at the University Research Conference by university management. It currently has 4478 records (3259 research records, and 1219 Virtual Herbarium collection records with images).

Outreach, Marketing and Faculty Engagement

Marketing and researcher engagement is undertaken on a needs basis, with the repository being embedded into institutional workflows.

For example, each researcher at USC has a staff web profile page. (E.g. see <http://www.usc.edu.au/University/AcademicFaculties/Science/Staff/022472.htm>). Each profile page has a 'Publications' heading which links to that researchers publications in the repository. Researchers know that to have their recent publications added and showcased on their profile pages, that the data has to be submitted to repository staff. When new researchers start at the University, Faculty Administrators obtain their research CV and forward this to repository staff to develop their publications listing in the repository to link to from their new profile pages.

Frequently asked questions have been created and are available on the Coast Research Database website.

In addition to research data, the repository infrastructure is used to showcase the USC Virtual Herbarium. This can be seen as a repository special collection. At present this collection includes 1200+ records and images of the physical herbarium collection. Planned growth is to expand this to include the complete Virtual Herbarium Collection of 8000+ records and images.

Relationships with Research Management Systems and other systems

Currently Coast Research Database does not have a self-submission component, and all research data goes through mediated entry by repository staff via XML uploads. This has ensured only consistent, clean data is present in Coast Research Database, while at the same time making upload as easy as possible for USC researchers (they simply have to email their new publication details or CVs to the research-repository email address).

The current model for HERDC and ERA reporting is as follows:

1. The ERA collection and HERDC collection is entered in Coast Research Database. Researchers can check their research and advise repository staff of any missing research publications or any alternative FoR allocations.
2. Repository staff then run a report from the repository pulling out only HERDC reportable data for that collection year.
3. This report is checked by Faculty Deans and Office of Research Staff and then ingested into the Research Master system.
4. The HERDC report is then run from the Research Master system.

This process enables the institution to collect and showcase 'all' research, but only include HERDC/ERA reported research in the official Office of Research 'Research Master' bank of data.

It ensures librarian-checked consistent data between the 2 systems, and enables the researchers to link to 'all' their research from their staff web profile pages to the repository.

Technical Environment and Information Technology strategies - Platforms/Architectures

Coast Research Database currently uses Fedora/Vital.

Repository staff have been very satisfied with the Fedora side of the repository, however have had problem after problem with the Vital component of the repository software. Due to such Vital stability issues (of which the indexing issue is the core issue), repository staff are still waiting to implement Handles and therefore have the repository harvested via various aggregators. The repository is however harvested via Google via the researcher profile page links (e.g. type **Dolphin Watch Tourism** into google. See second hit worldwide).

Staffing

Organisational Structure - Staffing and resources:

- Repository Staff report to Manager Library and Manager Office of Research
- HEW7/8 fulltime Research Information Systems Coordinator
- HEW5 20 hours per week Repository Officer
- HEW6 0.3 of an ITS Business Systems Analyst

Open Access

Currently only 1.6% of the repository content is open access fulltext, and the majority of this content are USC theses.

The focus on Coast Research Database to date has been on providing a complete showcase of USC research publications – and then integrating this data into institutional research reporting (ERA and HERDC).

It is hoped in the future there will be time to focus on increasing the open access fulltext content in the repository, however at present without researchers pressing for this, with tight staffing resources and with research reporting dates looming, the focus remains on providing a complete showcase of USC research and institutional reporting.

Please note however that while the fulltext open access content only sits at 1.6% of research data, the following steps have been taken to link a viewer to the fulltext of the research publication:

1. Every record is linked via 'Article Linker' link resolver to the Library holdings of fulltext subscription databases for USC students and staff.
2. Every record includes a 'publisher' URL/DOI leading to either a pay per view or free access fulltext version of the research.

Lessons learned and Future Plans

1 year:

- ERA repository readiness and reporting
- HERDC repository readiness and reporting
- Aim of 100% of USC research publications showcased in repository

2 years:

- ERA repository readiness and reporting
- HERDC repository readiness and reporting
- Aim of 100% of USC research publications showcased in repository
- Increase of fulltext open access content in repository

Advice:

- Working with the repository software has been the most difficult component of the USC repository experience. Select your software carefully.
- Collaboratively work with Office of Research wherever possible.

Sydney eScholarship Repository. Case Study

Sten Christensen – Sydney eScholarship Repository Coordinator

Sydney eScholarship

University of Sydney Library

February 2009

Summary

The Sydney eScholarship Repository [<http://ses.library.usyd.edu.au>] is an initiative of the University of Sydney Library and its commitment to support the research activities of the University of Sydney by providing an Open Access platform for research communities to safely store and access documents, publications and data. The University Library manages two repositories, one for Open Access material and one for material directly related for HERDC¹ and the upcoming ERA² exercise. The latter is a dark archive due to the copyright restriction of the stored material.

Institutional Overview

The University of Sydney is Australia's first University established in 1850. In 2008 the University had 46054 students enrolled; 3100 fulltime equivalent academic staff and 3014 postgraduate students³. The University has a strong emphasis on quality teaching and being a national leader in research.

The Sydney eScholarship Repository as well as the ERA/HERDC Repository are both managed by the Library. The latter is a successful collaboration between the University of Sydney Research Office and the Library in managing research outputs. The system that was developed offers independence for both to use their own systems.

The Sydney eScholarship unit was a major partner in the now defunct Australia Partnership for Sustainable Repositories and a collaborator as well as leader on a number its projects. Moreover the Sydney eScholarship and its Repository has collaborated on a number of digital projects across the University.

The Sydney eScholarship Repository was setup in July 2005 with the establishment of a trial community and collection from the Linguistics Department. The Repository was formally launched in May 2006 by the Vice Chancellor Gavin Brown. In choosing a name for the Repository the Library ran a contest to name it. Upon reviewing all the entries it was decided to go with the name originally given to the prototype namely the Sydney eScholarship Repository. As a tip for anyone wanting to name a service such as this; go with your gut instincts and call a "spade a spade".

The Sydney eScholarship Repository service is provided by the University of Sydney Library within the general policy and regulatory environment of the University of Sydney, particularly in relation to intellectual property policy, obligations within copyright legislation and requirements, Library policies, other general policies relating to theses regulations, Research Office requirements, Archive and Records Management requirements, and college regulations.

¹ HERDC The Higher Education Research Data Collection, see http://www.dest.gov.au/sectors/research_sector/online_forms_services/higher_education_research_data_collection.htm

² ERA Excellence in Research for Australia (ERA) initiative, see <http://www.arc.gov.au/era/default.htm>

³ Planning and Statistics University of Sydney, "Planning and Statistics," <http://www.planning.usyd.edu.au/statistics/>.

Its content parameters and guidelines are:

- The work must be produced, contributed, sponsored by or associated with a University faculty, department, school or centre
- The work must be scholarly or research oriented
- The work should not be ephemeral
- Material deposited should primarily consist of academic and postgraduate research material, with the exception of Honours theses or exceptions made by the department
- Not include any administrative records
- The work should be complete and ready for viewing at the time of submission.
- The work must be contributed in digital form. Preferred formats are those supported by the repository platform
- The author/owner should be willing and able to grant Sydney eScholarship the right to preserve and distribute the work via the institutional repository.
- If a work is part of a larger series or set of related works, these other works should also be contributed so that as full a set as possible is offered.

Relationships with Research Management Systems and other systems

As noted above the Library is managing two repository instances. For the purposes of the ERA requirements the model that best fits those described in the ARROW HERDC Working Group Interim report⁴ is Model 2: Research management system to institutional repository. All reporting and use of metrics are handled by the Research Office RIMS (Research Information Management System). The system was developed specifically for this purpose and to date is successfully managing 29000+ full text items. The Research Office RIMS handles the initial submission, metadata and reporting of the research outputs. It should be noted that there is much more metadata required for ERA reporting which is not necessarily important for Open Access archiving purposes.

Once the research output has been entered, verified and the quality of the object checked for readability, the RIMS generates a Submission Information Package that is sent to a shared area where the Library system grabs it and ingests it into the Repository. On successful ingest into the Repository, a handle (persistent identifier) is generated for the item. To allow look up and reporting through the RIMS, part of the reference to the full bit stream is sent to a common file (import.map) located in the shared area. The RIMS completes the reference to construct a url to the individual objects within the submitted item⁵.

The Sydney eScholarship Repository

Apart from the ERA/ HERDC repository the University of Sydney Library via Sydney eScholarship manages an Open Access repository, namely the Sydney eScholarship Repository. The focus of this repository is to provide an area where researchers and academics from or associated with the University can safely archive expose and disseminate their outputs. The Sydney eScholarship repository is archiving and disseminating:

- Entire sets of conference proceedings
- Image collections including the masters files of works
- Audio and Video of quest lectures
- Digital Theses

⁴ Arrow Working Group, "Arrow Herdc Working Group: Interim Report," (2008).

⁵ Gary Browne, Sten Christensen, and Venkatakrishnan Balasubramanian Appia, "From Research Management System to Digital Repository: Managing and Storing Research Outputs at the University of Sydney," in *eResearch Australasia 2008* (Melbourne2008).

- Research Outputs

Focus

Sydney eScholarship is made up of a number of innovative services that integrate the management of digital content with new forms of access and scholarly publication. The service consists of the Repository, Sydney University Press, SETIS - The Sydney Electronic Text and Image Service, and a Digital Project Analyst. There is a strong connection between the Repository and the Press and when possible the Press is archiving and providing Open Access to a number of its publications. This is also the case for a number of the departments on campus who via the repository are making available working paper series, many of which were never formally published.

The main aim is to preserve material which falls through the cracks, is not and will not be picked up by exercises such as the ERA yet still represent scholarship that needs to be exposed to the world.

Outreach, Marketing and Faculty Engagement

The outreach and marketing is focused on Faculty engagement. The core of the model is the Library's network of Faculty Liaison Librarians (FLLs), and the Library is in the process of formally defining the roles of this group in relation to supporting the Repository. Since the Repository's launch some roles have become apparent such as the need to market/sell the notion of Open Access and the Repository to faculties as well as the need for provision of various levels of support for the project process.

Being dependant upon a network of staff who work closely with the faculties requires an extensive level of support. This varies depending upon the complexity of the material or how a department or centre wants their collection managed and ranges from identifying potential material for archiving, setting up collections, managing copyright issues, providing statistics and metadata analysis.

The main advantage of having FLLs engage with the faculties about the Repository is that they have already developed a level of trust, know their clients and expectations and know the types of material that is available. This local knowledge is invaluable at the project scoping stage and helpful in maintaining the faculty's interest in the Repository. It is also important for the sustainability of the service.

Projects are initiated through the FLLs, though some come directly to the Repository Coordinator who refers them back to the relevant FLL. The basic work flow involves an analysis of the proposed material to see whether or not the Repository is the most appropriate place for it. If it is, a project scope is done and depending upon the complexity of the setup a project plan is created. In many cases a straight collection only warrants the soliciting of details from the department including information about copyright and some training. Once the collection is set up the department manages the submission and approval of material.

The Sydney eScholarship Repository uses a copyright release form developed by the University's Office of General Counsel (OGC). This form can either be signed or clicked through as part of the repository submission process. As part of its move toward an Open Access policy the University will set up user generated author amendment forms drafted by the OGC to publishing licences ensuring that Open Access publishing of research outputs. This will be based on the MIT Scholarly Publications module.⁶

⁶ MIT Scholarly Publications Module <http://info-libraries.mit.edu/scholarly/mit-copyright-admendment-form/>

Copyright releases are always a vexed point and one of the main barriers to moving forward with Open Access archiving. This is in part due to author ignorance, publishers Open Access policies, risk adverse solicitors and general confusion as non legal people try to deal with a piece of legislation.

Apart from copyright issues and fear of infringement, which is sometimes unfounded, significant barriers holding back the Repository's expansion include the :

- Lack of resources to fully develop the tools and support infrastructure that assist with the submission of material. An overall lack of integrated tools to easily submit material from the desktop makes it very difficult to not only capture material but also ensure continued deposition of materia by the end users. There are documented resources available such as ICE⁷ and Sword⁸ but these integration with the Repository.
- Even with a strong 'pitch' and raft of data and examples it is difficult to penetrate centuries of established academic publishing practice. Knowledge about and a trust in Open Access publishing and archiving remain an issue and is holding back expansion. Many researchers are not sure how Open Access works, fear plagiarism and loss of control of their work.

Technical Environment and Information Technology strategies - Platforms/Architectures

The Sydney eScholarship Repository and the ERA/HERDC Repository use DSpace version 1.5 as their operational platform. The Library chose DSpace because it wanted an open source solution which came with a user interface and had the resources to support it. At the database layer the Repository uses PostgreSQL run on RedHat Linux 5.x. The set up for both repositories is backed up to tape and stored offsite by the University's central IT Service.

We have only made minor modifications to DSpace to suit the University of Sydney Library's environment. The philosophy is that by having less dependencies we have a more robust system. The DSpace user community as a whole is very active and any patches required are easily available. .

DSpace comes with support for qualified Dublin Core and as discovery metadata suits our clients' needs. It is possible and relatively easy to incorporate other flat structure metadata schema. DSpace out of the box is OAI PMH compliant and our metadata is harvested by numerous services including Google, Google Scholar, Thomson's Web of Science Web Citation Index, OAlster , the National Discovery Service as well as the ADT. It should be noted that for the latter we have put in place a specific metadata crosswalk to comply with the ADT requirements.

Sustainability

The Sydney eScholarship Repository service is staffed by the Repository Coordinator and this position works in collaboration with the activities of the unit's Digital Analyst. IT support for the Repository comes from the Library's IT Services unit with much of the work in marketing the service done by the Faculty Liaison Librarians. This is an important support base especially in sustaining the

⁷ Australian Digital Futures Institute, "Integrated Content Environment (Ice)," <http://ice.usq.edu.au/>

⁸ Joint Information Systems Committee (JISC), "Simple Web-Service Offering Repository Deposit," <http://www.swordapp.org/>

service in a tight resource environment. It should be noted that the services of Sydney eScholarship are currently funded by the Library up until the end of 2009.

The Sydney eScholarship Repository Coordinator reports to the Director of Sydney eScholarship who in turn reports to the University Librarian. In relation to ERA activities while the Library manages the storage of the data, IT support reports to the Repository Coordinator as well as the Manager of Information Systems, Research Office.

The model for management and sustainability is to push the addition of metadata and loading of material onto the authors or their agents, with instructions from the Sydney eRepository Coordinator. The forms for capturing the metadata are set up specifically for each collection and this enables some quality control. Overall the quality of the metadata is high though other than regulating what can be captured and having workflows in place for administering what is added to a collection there is no formal quality control.

Metadata for the ERA/HERDC repository is handled by the RIMS and follows strict quality control. The metadata coming from the RIMS, however it is heavily qualified so is not suitable for OAI harvesting. The aim was to replicate the data stored by the current RIMS. This way of working will change with the implementation of a new RIMS in mid-2009.

Future Plans

Long-term planning for this area is always fraught with challenges especially in the current economic climate, the ever- evolving world of the ERA and the rate of technological change. While preservation and stability are at the core of what the Sydney eScholarship Repository offers, the maximum period that we plan for is two years and even then the plans change due to the above mentioned factors. For the next one to two years (2009 -2010) we are ambitiously planning to:

- Define the services of the Sydney eScholarship Repository amongst the plethora of data and digital services as well as other initiatives on offer within the University and the tertiary sector. The Sydney eScholarship Repository will seek to fully define its market, core services and the capacity to deliver those core services. This will involve a review of the services, working with clients to establish what we are successfully able to provide with the resources and constraints.
- Secondly the Repository needs to integrate its asset management capability with the Sydney Electronic Text and Image Service (SETIS) in storing this unit's data. SETIS is currently being redeveloped. The Repository needs to focus on the integration of DSpace with open source delivery tools in particular XTF to potentially deliver output.
- Thirdly the Library is working with the University to develop and implement an Open Access policy by advising it on all aspects of the process to assist in the development of support systems (Copyright IP), provide a mechanism to capture or acquire the data and metadata, and ensure that the infrastructure and platform (DSpace) can successfully manage such an endeavour. To this end we see the HERDC/ERA as a mandate for a specific set of data and as such we do not need to set another mandate as this maybe counterproductive. The focus for the University is to have the tools and the infrastructure in place to enable a smooth and easy transition for archiving. It should be noted that through the ERA/HERDC Repository we can potentially, with the change of a few DSpace parameters, make 30000+ items available. However to do this we will need to ensure that the rights and publisher requirements are all in place before doing so.

- Finally, working with our Faculty Liaison Librarians (FLLs) we will seek to develop strategies to acquire that data, which is born digital or being converted, that falls outside the narrow and targeted ERA/HERDC focus but is also relevant and should be archived and exposed. These include working paper series, image collections, grey literature and special quest lecture series. The media for this type of material varies and we are encouraging the different departments to store their archival copies as well as their presentation files.

Issues, Challenges and Lesson Learned

Some of the issues and challenges that have been easy to deal with have been mainly technological in nature. Both Repository platforms are run on DSpace which open source is giving us the freedom to move quickly to address issues and work out solutions either by ourselves or with the support of the DSpace community. Some of the technological issues we encountered include the development of the system for moving data and object from the RIMS to the repository, a batch loading capability, developing OAI cross walks and managing storage.

Managing DSpace has been relatively easy. As a system DSpace is very robust, relatively easy to set up and as it comes with a Web UI it allows quickly to get started with acquisition of content.

On the flip side there are number of things that we have found challenging such as cultural change and the persistence of the old publishing model preconceptions. This for most part, but not always, is a generational. We hope is that the newer generation of researchers coming through the system will be more comfortable with the concepts and processes of Open Access archiving through their exposure to some of Web 2.0 social interaction technologies such as Wikipedia, Flickr and Facebook. We work on highlighting the benefits of expending the effort to archive material. This is especially difficult in a “time-poor” environment, so a good narrative and a war chest of solid examples is essential. Related to cultural change is the notion of encouraging sustainable formats. This is not the “sexy” side of repository management but a core fundamental. Trying to instil and encourage good digital practice is difficult as it requires effort, time and in many cases more resources.

Another challenge has been to develop and implement the tools to simplify the process of getting researchers to place material in repositories. This includes a whole raft of things such as generating rights/ licences and author addenda to be included in agreements, single sign-on authentication for acquiring information about the author for licence purposes, integrated deposit tools such as those proposed by ICE (Integrated Content Environment) and the Digital Scholars Workbench (DSWB). Sydney eScholarship is taking a pragmatic approach and sees great value exploring the use of tools to acquire content. The “mandate” may not be enough, may increase resistance or, worse, encourage behaviour of providing the bare minimum. To quote the late Kerry Parker when speaking on tax in front of the 1991 Print Media Inquiry;

“... I am not evading tax in any way, shape or form. Of course I am minimizing my tax. Anybody in this country who does not minimize his tax wants his head read ...”⁹

When academics or researchers are forced to give, especially in the “time-poor” economy, we will get the minimum and this is not what we see as the end game. Changing the culture is as important as acquiring the data and it should lead to a greater flow.

⁹ Paul Barry, *The Rise and Rise of Kerry Packer Uncut*, Rev. ed., 070.5092 20 A (Milsons Point, N.S.W.: Transworld Australia, 2007), 410.

It is important to make researchers, see things beyond the ERA. As noted above one of the main ambitions for the Sydney eScholarship Repository for the next two years is to archive and expose the materials that fall outside the ERA/HERDC exercises. This poses an interesting challenge for the notion that capturing for the ERA/HERDC covers most outputs, is not true. The Government has provided money for the development of systems to manage this data and this has, in some cases blinkered the vision of some repository services and created a narrow collection focus.

Finally the last issue is linked to breaking the aversion to risk in relation to copyright and demystifying the urban myths around it. Managing risk is a major part of a Repository Manager's role. One of the more difficult aspects of the job has been trying to overcome the aversion to risk on the part of the University's legal office. The misconception seems to stem from the view of protecting IP and not seeing the value of Open Access in terms of increasing exposure to the research outputs of the University. This approach is compounded by overly complex and convoluted Open Access policies on the part of publishers and general misconception about rights.

What advice would you give others?

To end this case study there are five pieces of advice we can convey:

- Don't wait for the mandate, for it may not be the panacea that you think, and if you do you receive the mandate you must have the tools in place otherwise you will lose the client's trust and possibly the momentum. The Trusted Repository is still at the core of our work. Working to change the culture or practices at an institution let alone a whole sector takes time.
- Trust your instincts. There's plenty of advice out there but seek your counsel wisely. Other repository managers are generally willing to share their knowledge and the inside story.
- Be prepared to manage risk especially with regards to copyright and IP. Work out what is myth, smokescreen and real risk.
- Develop good narratives of your successes¹⁰. Success stories are always a good selling point and can persuade others who might be sitting on the fence to come over to your view.
- Don't lose the trust of your clients by promising things that you can not deliver. Loss of reputation can be very damaging.

¹⁰ Here I am indebted to Rowan Brownlee – Digital Project Analyst from Sydney eScholarship for these thoughts and his wise comments on this matter.

Bibliography

(JISC), Joint Information Systems Committee. "Simple Web-Service Offering Repository Deposit."

<http://www.swordapp.org/>.

Arrow Working Group. "Arrow **HERDC** Working Group: Interim Report." 2008.

Paul Barry. *The Rise and Rise of Kerry Packer Uncut*. Rev. ed, 070.5092 20 A. Milsons Point, N.S.W.: Transworld Australia, 2007.

Gary Browne, Sten Christensen, and Venkatakrisnan Balasubramanian Appia. "From Research Management System to Digital Repository: Managing and Storing Research Outputs at the University of Sydney." In *eResearch Australasia 2008 Melbourne*, 2008.

Institute, Australian Digital Futures. "Integrated Content Environment (Ice)." <http://ice.usq.edu.au/>.

Dorothea Salo "Innkeeper at the Roach Motel." <http://digital.library.wisc.edu/1793/22088>

University of Sydney, Planning and Statistics. "Planning and Statistics."

<http://www.planning.usyd.edu.au/statistics/>.

1. Please give us the following details • Name of Institution • Name of the repository(ies) and web address(s) • Author(s) of the case study (include contact details if possible) • Summary/abstract • Date prepared

University of Western Sydney Research

Repository <http://arrow.uws.edu.au:8080/vital/access/manager/Index>

Robyn Benjamin r.benjamin@uws.edu.au

Euwe Ermita e.ermita@uws.edu.au

Lisa Tyson l.tyson@uws.edu.au

Michael Gonzalez m.gonzalez@uws.edu.au

23 April 2009

2. Institutional Overview Include information about your institution, size, focus, FTE, culture • Where the repository is managed from • Partnerships, collaborations • Name of the repository – any tips for others on choosing a name • Repository timeline • Mission statements • Regulatory environment within your institution

The UWS Research Repository is managed by UWS Library staff, working closely with staff from the Office of Research Services. It was decided not to "name" the repository as such names are often catchy at the time of implementation but lose meaning over time. Work commenced on populating the repository in 2007 and is ongoing. Currently, research outputs are forwarded to the Library from ORS for inclusion in the repository. Through 2009, Library systems staff are working on implementing self submission integrating with the inhouse ORS research database.

3. Comment on • Your institutional model based on the Arrow HERDC report • <http://www.arrow.edu.au/docs/files/arrow-herdc-interimreport-june08.pdf> • Institutional embedding and innovative practices

UWS is working toward Model 3 – shared input – for self-submission implementation. In late 2008, work on populating the Repository was mainstreamed within the responsibilities of the Cataloguing and Metadata Team. Library systems staff work closely with this team and ORS to ensure the best possible workflows.

4. Outreach, Marketing and Faculty Engagement Give details about any of the following • Strategies and Services • Support required • Workflow • Frequently asked questions • Copyright, licencing,(include use of Creative Commons) and legal aspects • Barriers to moving forward

Research Services Librarian liaises with Faculty and PG students to raise awareness of the repository and encourage input. The University Librarian is a member of the UWS Research Committee which also assists in raising the profile of the work undertaken within the library in relation to the repository. Wherever possible, within Copyright regulations, authenticated access to full text is made available.

5. Relationships with Research Management Systems and other systems • Interoperability, data exchanges, and ongoing relationships • Data duplication and/or matching • Single

signon and authentication issues • Output to personal web pages, CV creation, Bibliographies and Social Networking tools

Extraction from the UWS research management system is currently a manual process. Data is 'cleaned' by library staff and metadata is added prior to batch load to the repository (undertaken by library systems staff). Prior to data being forwarded to the library by ORS, deduplication is undertaken. In the future, records from the repository will be utilised for UWS PG and academic staff profiles.

6. Research Assessment and Reporting Comment on • How your repository supports formal research assessment exercises • Other reporting exercises generally • Discuss on metrics etc outside of these formal exercises

In relation to the ERA, the repository will be the authoritative source of data.

7. Technical Environment and Information Technology strategies -Platforms/Architectures • Software Platform(s) • Database and Operating Systems • Unique setups • Mirrored Systems, Redundancy and backup strategies • Customisations – how heavily is your system customised? • Metadata schemas • Open URL linking • Harvesting

• Software platform = VITAL (VTLS) • MySQL and Solaris 10 • Zoned on a T2000 server • File system hosted on a redundant SAN • Customised to meet UWS needs (cf look and feel, metadata ...) • MARCXML and Dublin Core • OAI-PMH harvested by National Discovery Service and ADT

8. Sustainability • Organisational Structure - Staffing and resources • Funding and external grants • Management structure – reporting lines • Rate the quality of your metadata • How much Open Access full text does your repository contain? – is it a question of resources or copyright?

As noted above, work on the repository is now mainstreamed and as such sustainable in the long term. ASHER funding is received from ORS for a temporary, part time position to specifically assist with preparation for the ERA. Associate Librarian, Corporate Services has overall responsibility for management of the Repository. As metadata is created by library staff (rather than academic) we are confident of its quality and consistency.

9. Lessons learned and Future Plans • What is next? 1 year? 2 years? Longterm • What has been easy to solve? • What has been difficult to solve? • What advice would you give others

Select a robust and proven platform and infrastructure. Critically, population of the repository needs to be mainstreamed as early as possible. It needs to be seen and accepted as part of regular/standard responsibilities rather than "special" or added on.

Educause 2009 – Repository Case Study

Victoria University
Jane Miller
Electronic Information Systems & Services Librarian

Name of the repository(ies) and web address(s)

Victoria University Institutional Repository (VUIR)
<http://eprints.vu.edu.au>

Victoria University Research Online Publications Store (restricted access)
<http://vurqf.vu.edu.au>

Items in the Repository: 1232 – approximately 65% fulltext – 35% metadata only.

Institutional Overview

Victoria University is a dual sector University based in the western suburbs of Melbourne. The University has 11 locations and approximately 47,000 on shore and offshore students across both the Higher Education and Vocational Education sectors.

Victoria University has had an open access repository in place since the end of 2005. The repository was initially implemented by the Library and is currently managed within the Information Resources branch of the Library with the recently established position of Research Librarian undertaking promotional and awareness raising activities.

Models

The ARROW model that best describes the current repository process is model 2. We are currently working closely with the University's Office for Research on ways that the repository can integrate with the research system to support publication collection processes associated with HERDC and ERA reporting.

Faculties currently enter their publication information in the research system and this year we have begun to establish processes by which the repositories can take feeds of this data and attach copies of publications. We then supply a feed back to the research system of the location of the publication itself.

The University has recently made submission of copies of the University scholarly and research publications to the repository mandatory (as permitted by copyright/publisher permission). This is a new policy that was confirmed in early 2009. The challenge is to embed the compliance to the policy as part of the scholarly publications cycle.

Outreach, Marketing and Faculty Engagement

The Research Librarian in collaboration with Faculty Liaison Librarians is largely responsible for the marketing and promotion of the repositories and their benefits. Much of this educational process is done in conjunction with provision of information about the range of services and resources the Library is able to provide to researchers and is done via workshops, attendance at meetings and general day to day engagement.

Due to the need to adapt the focus in 2007 to the, then, Research Quality Framework, the repositories have been experiencing a period of transition in relation workflow, staffing and procedures. Over the last two years, the role and nature of the repository and the establishment of the second instance has meant the service and system is evolutionary in nature.

Much of the entry of metadata and submission is done by staff in Cataloguing. VU does not make self submission unavailable but it is not the preferred method of the majority of academics who contribute content.

The process of submission is usually initiated by an academic providing the Research Librarian with a copy of their publications list and the Library then checking how many of these can be included on the repository . OAKList and Sherpa are used extensively by the Cataloguing staff to ascertain publisher permission status and in the event that they are permitted to be archived, the author is requested to provide a copy of their final version (post print) and a form that indicates they give authorisation to have it included on VUIR and that it has peer review or the permission of their Head of School for inclusion.

In the event that OAKList or Sherpa have no information about the publisher or indicate institutional archiving is not permitted, the author is contacted and informed of this. It is then up to the author to pursue the matter with the publisher should they wish to proceed further.

The Library was initially, undertaking additional work on the seeking of publisher permission but it became difficult to manage. Unfortunately this has meant that the author is required to undertake this pursuit of permission.

The Library is planning to work with the University on providing more guidance to author's at the commencement of the scholarly publishing cycle in terms of the agreements they sign in regard to the intellectual property associated with publications.

Relationships with Research Management Systems and other systems

The Library and Office for Research are currently working on the process for data exchange and to date, this has largely been achieved via simple XML feeds for output and input.

There is no formal process for output to web pages for online CVs but the nature of the platform means that authors are able to get a persistent link to their own author page listing all publications.

Technical Environment and Information Technology strategies -Platforms/Architectures

Both repositories are located on the University of Southampton's GNU Eprints platform. Two repository instances have been established on one dedicated server. There has not been extensive customisation undertaken to the system but we have added modules from University of Queensland and University of Tasmania to enable us to provide usage statistics to authors.

The VUIR repository is harvested by the Arrow Discovery Service as well as OAIster and Australian Digital Theses. We have experienced some issues with harvesting which are still to be resolved.

Sustainability

To date the repository staffing has largely been a mix of staff time from existing positions and teams. The Electronic Information Systems & Services team were the initial implementers of the technical infrastructure and worked with Information Technology Services to establish the repositories.

As time as passed, Cataloguing staff have become increasingly involved with the workflows associated with the open repository and actually enter the metadata and manage the submission process on behalf of submitting authors.

The addition of the role of Research Librarian has meant that the promotion and provision of end user instruction/information is been taken up by this position but there is still an ongoing need to establish some sort of dedicated staffing at the management end of the repository to ensure workflows are established and the repository becomes a mainstream service with staff resources to support its growth.

The efforts that the Library has put in working with the Office for Research on the ERA and HERDC projects, has allowed the Electronic Information Systems & Services Team to develop methods for exchanging data but reinforced the need to put in place a position with an overview of the various aspects of the repositories and associated workflows and to bring together the various staff and processes from across the Library to enable a cohesive service to facilitate the further development of the Repository. The repository needs a Repository Coordinator or Manager in order to ensure it continues to develop in a considered and coordinated way.

The Library is currently exploring opportunities for implementing such a position in order to support the ERA and HERDC projects. This new position will report to the Electronic Information Systems and Services area but maintain close links with Cataloguing, the Research Librarian and Office for Research.

Technical support for the platform and hardware is provided by University Information Technology Services and while this has worked relatively well, they have had to develop their own knowledge of the Eprints application. This, combined with their other competing duties, has meant that there have often been delays in implementing significant hardware and software changes.

Now that the platform is relatively established, one of the staff in the Electronic Information Systems & Services team is responsible for ongoing customisation and liaising with Information Technology in the event that assistance is required.

12B Lessons learned and Future Plans

While the platforms are relatively inexpensive to install, there are many more issues in the process of establishing an institutional repository than is initially apparent. The original emphasis of many repository projects and presentations in the early period, concentrated on the benefits for academics of archiving post prints on an institutional server. Since those early days, issues around how to sustain these repositories and staff them have become increasingly apparent.

This is also true of the labyrinth that is Copyright and publisher permissions and knowing how much further individual Libraries can pursue issues of Copyright apart from the information provided by OAKList and Sherpa has been a stop start process.

The lesson we learned is that requesting copyright permission is staff intensive and while desirable for the submitting author, not something the Library could sustain on a one to one basis.

The other challenge has been related to devoting appropriate staff resources to the Repository. In the beginning it was a niche service and staff from Cataloguing and Electronic Information Systems and Services were undertaking work on it in addition to normal duties. As the repository develops and becomes part of the University's policy context and plays a role in the Research evidence collection process, there is a need to dedicate staff to its maintenance and development.

For VU Library, this is the next major area to be addressed as is the establishment of regularised and effective workflows to ensure its growth and sustainability.

Research Repository Case Study

ScholarlyCommons@AUT

Craig Murdoch, Digital Services Coordinator, AUT University Library

March 2009

1. Please give us the following details

- AUT University
- ScholarlyCommons@AUT <http://aut.researchgateway.ac.nz/>
- Craig Murdoch, Digital Services Coordinator, craig.murdoch@aut.ac.nz
- 18 March 2009

Abstract: This case study describes the development and first three years of ScholarlyCommons@AUT, the research repository of AUT University. It covers the institutional context; cooperation and integration across University departments; the technical environment; resourcing and constraints; challenges faced and solutions implemented.

2. Institutional and Repository Overview

AUT University is New Zealand's newest university, established as a University in 2000, after 105 years of educating people in the Auckland region. The University takes pride in engaging with the community, undertaking exciting and meaningful research, offering a unique style of teaching and research, and in particular, watching graduates flourish in their chosen careers. AUT University currently has 16500 FTEs, 4 faculties and 16 Research Institutes.

The research repository, *ScholarlyCommons@AUT Putunga Rangahau* is managed from the University Library (the Māori name was suggested by John Moorfield, Professor in Māori Innovation and Development at AUT University and conveys the idea of a storage place for research). ScholarlyCommons@AUT was initially established in 2006 on the Digital Commons platform, at that time offered by ProQuest. The repository was migrated to DSpace in April 2008. DSpace was selected by the Library Consortia of New Zealand (LCoNZ - <http://www.lconz.ac.nz/>) as the platform for a hosted repository solution to be used by three of the LCoNZ partner libraries - AUT University, University of Waikato and University of Otago.

DSpace is hosted by University of Waikato, Information and Technology Services Division and LCoNZ employs a dedicated Institutional Research Repository Administrator/ Developer working full time out of the Waikato ITS Division. This position manages all of the technical requirements of DSpace as well as offering the following:

- development of customised solutions,
- liaison with the wider DSpace community including submission of patches and enhancements,
- assistance with customisation of the DSpace web interface,
- assistance with content loading.

As at March 2009 ScholarlyCommons@AUT contains 380 full text theses.

3. Integration with Research Management System

The University is currently in the process of selecting and implementing a new research

management system (RMS). The Library has worked closely with the University Research Office over the last 12 months to ensure that ScholarlyCommons@AUT is considered as an integral part of this planning. It has been agreed that the University will operate according to institutional Model 2 in the Arrow HERDC report. This is for the following reasons:

- the RMS will be the central collection point for all University research outputs because of the need to focus on Performance Based Research Funding reporting,
- ScholarlyCommons@AUT will contain only full text, open access objects, and will thus be a subset of the content in the RMS,
- while researchers will be encouraged to submit digital copies of their research to the RMS, the Library will check copyright and source the most appropriate version of the research for the repository,
- the Library will bulk load all content to ScholarlyCommons@AUT (no self-submission) to reduce data entry requirements for researchers and to improve quality of metadata.

At the present time the Library receives monthly CSV feeds from the interim RMS. In the future it is hoped to improve the automation of this process.

Deposit of Doctoral Theses and Masters Theses and Dissertations is mandated by the University. These are self-submitted with graduation not possible until the DSpace email receipt has been forwarded to the Postgraduate Centre. Library staff check metadata and publish the thesis or dissertation.

4. Outreach, Marketing and Faculty Engagement

The Library has spent the past year working on a project to make staff research outputs available on ScholarlyCommons@AUT and the first of these will be live in April 2009. Although this may seem like a considerable time, it has been time well spent for several important reasons:

- integrating the repository into the workflow, planning and mindset of the University Research Office, enabling steady and regular delivery of metadata and objects for use in the repository,
- communicating both formally and informally the value and potential of the repository to faculties and researchers,
- gaining an understanding of the issues surrounding copyright for objects which have already been published and developing workflows to integrate this knowledge,
- developing workflows which enable the Library to deliver an effective repository solution within our resourcing constraints,
- developing policies and gaining approval through appropriate forums,
- nurturing champions at faculty and Research Office level.

The most significant barrier to growing content continues to be publishers' copyright restrictions and the related problem of efficiently sourcing permissible versions of research outputs. The hope is that researchers who may be reluctant to provide objects or who do not see the value of the repository will be convinced as content grows and usage statistics show the reach of the repository.

5. Relationships with Research Management Systems and other systems

Please see above (section 3) for a description of work thus far with regards to the University research management system.

ScholarlyCommons@AUT uses standard DSpace authentication. Authentication is only required for submission of theses by postgraduate students. Students can self-register

(submission is set to anonymous - anyone can submit) - this eliminates any delay in the submission process. As all objects in ScholarlyCommons@AUT are open access and all content except these is bulk loaded there is no other requirement for authentication.

The Research Committee plans to introduce Academic Portfolio software and the intention is to work towards integration with DSpace.

6. Research Assessment and Reporting

At this stage ScholarlyCommons@AUT does not formally support Performance Based Research Funding reporting or assessment. With the next round of PBRF not due until 2012 the Research Office is concentrating on implementation of the new RMS, after which closer integration with ScholarlyCommons@AUT will be considered.

Because of dissatisfaction with the standard DSpace statistics, the LCoNZ IRR Administrator/Developer has developed a standalone statistics package, based on standard log analysis and which requires no patches to DSpace java source or alterations to database schema. This is used to provide online, unmediated access via ScholarlyCommons@AUT to usage statistics and reports. Faculties and Research Institutes have shown interest in this as the repository moves into hosting staff research outputs.

Usage reporting is also provided via Google Analytics, which gives a broader and more interesting perspective for a wider community. (See http://www.aut.ac.nz/library/library_resources/scholarlycommons/about.htm for monthly reports.)

7. Technical Environment and Information Technology strategies

The DSpace web interface has been minimally customised for ScholarlyCommons@AUT. Most customisation has been focused on branding, colours and text labels. ScholarlyCommons@AUT runs on the original jspui DSpace interface. Only minor consideration has been given to the more flexible xmlui (manakin) interface, simply because the jspui interface works very well.

The following patches and add-ons have been created and either submitted to DSpace or shared with the community:

- RSS date handling
- improved DOI (digital object identifier) handling
- Māori macron search/browse filters
- standalone statistics package
- auto-notify script for deletions (to notify harvesters)

The metadata schema used is Qualified Dublin Core. This is supplemented for theses with the 'thesis' elements of "ETD-MS: an Interoperability Metadata Standard for Electronic Theses and Dissertations" (<http://www.ndltd.org/standards/metadata/etd-ms-v1.00-rev2.html>). There are also a small number of custom metadata fields which collect responses from self-submitters, and which form no part of the brief item display on ScholarlyCommons@AUT.

ScholarlyCommons@AUT content is exposed for harvesting using the OAI-PMH standard for metadata harvesting. 32 per cent of repository traffic originates from referring sites with harvesters such as Google Scholar, Australasian Digital Theses and the Kiwi Research Information Service strongly featured. 62 per cent of repository traffic originates from search engines, almost all from Google.

ScholarlyCommons@AUT runs on DSpace 1.4.2 with an upgrade to 1.5.1 planned for Easter 2009. Operating System is Linux and database is PostgreSQL 8.1. We are investigating Slony-I for database replication/failover. There are 7 LCoNZ DSpace servers - production and test for each library, and a general test instance. Production servers are backed up daily via Legato NetWorker with one full backup every weekend. Test servers are backed up daily via Amanda.

8. Sustainability

The decision to share hosting and the costs of an Administrator/Developer was a very good one. The AUT Library would not have been able to contemplate such a position alone, and the benefits have been excellent (though this is admittedly somewhat dependant on the person in the position). These have included:

- handing over what the Library considers to be non-core business - essentially the technical side - allowing the Library to focus on content and promotion,
- increased exposure in and contribution to the DSpace community,
- increased cooperation and collaboration within LCoNZ,
- the development of increased expertise within LCoNZ and the individual Libraries.

There are limited resources available within the Library to work on the repository. Less than 1 FTE is dedicated to the repository, with this being spread across multiple positions. This includes:

- Digital Services Coordinator - liaison with IRR Administrator/Developer, web interface customisations, DSpace administrator functions, content ingest of Research Outputs,
- Research and Postgraduate Coordinator - metadata and copyright checking, and approval of theses and dissertations, liaison with students and Postgraduate Centre,
- Digital Services Library Assistant - metadata and copyright checking, and approval of theses and dissertations,
- Information and Education Services Team Manager - promotion and communication,

with many other positions providing assistance with specific tasks.

The Library has also used staff on short term contracts for data entry and copyright checking.

ScholarlyCommons@AUT received initial funding from the University to implement and populate the Digital Commons repository but has received no further funding or external grants. Dividends from LCoNZ have been used to pay for contract staff. All other repository activity is integrated into the workload of existing Library staff.

The repository contains 100 per cent open access full text. The Library believes that ScholarlyCommons@AUT should be an open access full text repository. Theses which are embargoed or have copyright issues are not loaded into ScholarlyCommons@AUT. Only research outputs with full text available to be loaded will be accepted.

9. Lessons learned and Future Plans

The next major milestone for ScholarlyCommons@AUT will be going live with non-thesis research output collections in April 2009, and the ongoing development of these collections. This will involve a continual process of improving the data sharing between the Research Office and the repository, and looking for efficiencies in copyright clearance and version collection.

In the future the Library would like to consider hosting other types of research

collections - working papers, discussion papers, faculty seminars for example.

One of the interesting questions to be addressed is how the repository will effectively host non-text research objects. The University has a large and active Faculty of Design and Creative Technologies and their research output is of exceptional quality, of interest to a wide audience and, in many cases, non-traditional in form.

LCoNZ is actively investigating improvements in the self-submission process using SWORD, with a view to long term improvements in integration with research management systems.

The ScholarlyCommons@AUT experience has been typical in terms of problems and issues. Technical and metadata issues have been relatively easy to solve, whereas marketing and policy issues have been difficult to solve. What is clear is that external hosting and dedicated 'non-Library' technical support has been extremely valuable in terms of allowing Library staff to focus their efforts on the most important issues of content, communication, promotion and workflows. The most difficult, or at least time consuming, issue continues to be the question of publisher copyright requirements and sourcing appropriate versions of journal articles and conference papers.

Research Repository Case Studies: UC Research Repository

Institution: University of Canterbury, Christchurch, New Zealand

Repository: UC Research Repository

Address: <http://ir.canterbury.ac.nz>

Author: Grant Barrie, Institutional Repository

Coordinator, Grant.Barrie@canterbury.ac.nz

The UC Research Repository was launched in May of 2007. The Repository is an archive of scholarly research and postgraduate theses – more than 99% of which are openly accessible as full text records. Our repository is tightly integrated with the University's research management system and incorporates a set of custom-written statistical reports that enable authors and their managers to keep track of their publications, easily and efficiently.

While DSpace 1.4.2 is used as the repository container, all additional functionality and customisations are delivered via a suite of bespoke PHP programs. This decoupled approach to IR development ensures that future upgrades will be drama free, and leaves our data portable enough to even entertain a change in vendors – without the risk of losing any hard-won programming gains.

Date prepared: 27 March 2009

Institutional overview

The University of Canterbury, Te Whare Wānanga o Waitaha, is situated in Christchurch, the largest city in the South Island of New Zealand.

STUDENTS – as at 31 December 2008

Student Headcount: 22,973

International Students: 2,291

Domestic Students: 20,697

Undergraduate: 13,063 (88%)

Postgraduate: 1,714 (12%)

Total Taught EFTS: 14,860*

College of Arts: 3,335 (22%)

College of Business and Economics: 2,375 (16%)

College of Education: 2,306 (16%)

College of Engineering: 2,841 (19%)

College of Science: 2,478 (17%)

School of Law: 722 (5%)

Other: 804 (5%)

STAFF – as at 31 December 2008

Staff FTE: 1,619

Continuing Academic Staff FTE: 595

Continuing General Staff FTE: 1,024

FINANCE – as at 31 December 2007

Income: NZ\$257 million

The UC Research Repository is managed by the Library's IT department, with one half time position dedicated to the management of the Repository, and numerous other people involved throughout the library - Information Librarians solicit material from academics and verify the provenance of the outputs, an administrator checks the copyright status of work and cataloguers add subject headings to the records and check the metadata before publication.

The Repository is integrated with the University's Performance Based Review database (UC People & Research Database, UCPRd), which is managed by the University's Research Office. The integration and our close working relationship with the Research & Innovation team provides some very obvious and beneficial synergies for both parties.

IT support is provided within the Library IT department and by the University's general ICT Service. A high level of IT support and access to a programmer is key to the tight integration of the University Research database and the ongoing development of the Repository.

The Repository has been operational since May 2007.

Models

Our institutional model was developed around the general principle that data about research should be collected once, to be used in multiple contexts. Hence, you could say we broadly conform to the Arrow model 2, where the data is captured by our research management system, with the data then flowing to the institutional repository. Full text documents are appended to the Repository via the UCPRd, so in actual fact, academics never directly use DSpace.

Our model is not yet consistent - our theses travel via a separate workflow. However, we have begun work on creating a new workflow that brings greater uniformity to our procedures. In part, this is due to an existing weakness in the research management software, which we are attempting to address to the mutual benefit of ourselves and the Research & Innovation unit.

Outreach, Marketing and Faculty Engagement

The University Library has a simple and understated strategy of reaching out to academics through existing relationships between themselves and their subject specialist librarians. There are also considerable 'cross sell' opportunities presented by the training and guidance requirement around the recording of scholarly outputs by academics in the UCPRd.

Support for everyone involved, from academics to cataloguers, comes from the dedicated half-time person situated in Library IT. Hence difficult copyright or philosophical questions normally land on the lap of the IR Coordinator.

The workflow is carefully mapped and Repository responsibilities (such as cataloguing) have been incorporated into the job descriptions for all personnel. New staff are briefed on the workflow and their respective roles and responsibilities.

Relationships with Research Management Systems and other systems

The relationship with the research management system is tight, though not complete. Data flows between the UCPRd to the Repository - we get the metadata we require and return information about when an item was ingested and what the handle for the item is. However, once items are ingested into the Repository, changes to the metadata do not flow from one system to another. While this can be considered a weakness, records are only imported once the metadata has been fully verified, which lessens the potential for data to vary between systems.

In terms of reciprocation, content in the Repository is connected to the University's dynamically created researcher profile pages - so if an Engineer has articles stored in the repository, the bibliography on his profile page (drawn from the research management system) will have active links to the full-text document within the Repository.

Research Assessment and Reporting

While our Repository does not support the research assessment process, we do keep usage statistics for the Repository. The default statistics package in DSpace is disappointing, and in our installation at least, faulty. The Minho stats package is a nice attempt to provide more detailed statistics but still doesn't meet our immediate needs for monitoring use. The fundamental flaw – in our opinion – is the lack of authority control when it comes to authors. After much soul searching and prevarication (in the hope that the DSpace community would provide something better), we developed a statistics package that allows academics to monitor the use of their repository items.

The stats module makes a copy of the DSpace web logs and stores them in a separate database. Having a stats program outside of DSpace makes us fairly immune to the sort of upgrade concerns an integrated module might have, while giving us the all important authority control missing from DSpace itself. The metrics themselves are used by faculty managers and staff to make them feel happy about their work – and happy to keep contributing. Naturally, we do caution them on the reliability of web statistics in general.

Technical Environment and Information Technology strategies - Platforms/Architectures

The UC Research Repository currently uses DSpace, version 1.4.2. The supporting services are written in PHP (uploading records from the UCPRd, statistics, workflow analysis, licence management and other administrative tools).

As we use DSpace for very little other than serving documents, we have kept customisations to a minimum, which makes upgrades simple and, hopefully, pain-free.

Sustainability

The Repository sits with the Library, and is managed by the Library IT department. The Coordinator role is a 0.5 FTE position (the only direct salary cost directly attributable to the Repository), with support from subject-specialist librarians and Collection Services staff. The IR Coordinator reports to the Library IT Manager, with the over all Repository function falling under the responsibility of the Associate University Librarian.

The UC Research Repository is 99% open access full text. Copyright being the disabling beast it is, we have a large number of items in an uncompleted items workflow area, consisting of items that are either embargoed or not in compliance with copyright – in either case, unless a record is able to be released into the wild with its full-text document, we do not make it visible as metadata alone.

In terms of quality, we think the metadata we create and store is robust. Records are subjected to an agreed level of verification process before being ingested into the system, and are checked again as part of the final workflow step.

Lessons learned and Future Plans

Our immediate plans are focused on boosting the growth of the Repository – both in terms of item numbers and participants. The Repository is strongest in the Engineering and Sciences disciplines, and we need to attract academics from the social sciences and humanities.

We have a medium-term project to digitise and upload older theses (new theses are added automatically). Along the way, we will continue to try and identify areas for improvement, while refining our workflow processes and procedures.

We have been lucky to have had a strong collaborative relationship with two other New Zealand universities – Auckland University and Victoria University of Wellington. The ability to workshop ideas and approaches with other institutions has proven very valuable. However, the particular environment of Canterbury has dictated a different approach to systems integration, and the development of a bespoke statistics package (and many other reporting tools) has been challenging.

In terms of advice, we would probably suggest gathering as many *candid* collaborators as possible – and this holds true at every life stage of the Repository, not just at the project's inception. An adequate level of IT support is crucial, as 'out of box' DSpace is clunky and almost encourages inefficiencies. Developing additional functionality outside of DSpace (or any other vendor's product) seems a logical approach – particularly for institutions that have a need for integrated repositories. The last observation – I won't call it advice – is that repositories should never be thought of or sold as 'institutional' (except perhaps to funders!); the role of our repository is to support individual academics and students who are passionate about the knowledge they create, by making their original research accessible to all who wish to read it.

From the Beginning: Teresa Chitty

Page from the OARINZ Wiki

08 April 2009, at 05:11 PM

Teresa Chitty

*Lincoln University*¹

*ResearchArchive*² 12.12.2008

Contents

1. *Institutional Data - Type, size, staffing, budget, existing systems* (p 1)
2. *The Institutional Repository Project at Lincoln University* (p 1)
3. *Repository Content and Usage* (p 3)
4. *Successes* (p 4)
5. *Issues and Challenges* (p 4)
6. *The Future* (p 5)

1. Institutional Data - Type, size, staffing, budget, existing systems

- Lincoln University has approximately 2,700 students studying undergraduate, postgraduate, and research degrees.
- In 2007, the university employed over 600 FTE staff, including 30 full time equivalent employees who worked at the university library.
- The library's budget is around \$3 million, including \$1.2m for staffing, \$500,000 operating costs, and approx \$1.2m for serials subscriptions (2007).
- The library holdings in 2007 included 35,679 print and electronic serial titles, and over 100,000 non-serial holdings (including 43,397 e-books).
- The library uses the Voyager integrated library system

2. The Institutional Repository Project at Lincoln University

- Initially, representatives from the Lincoln University library were in discussions with University of Auckland and University of Canterbury to be part of a joint Institutional Repository project, but there was a last minute rush and the proposal went ahead without Lincoln.
- At the same time, representatives from the OARiNZ^(w) project contacted the Lincoln University library, asking if they wanted to be involved in the

¹ www.lincoln.ac.nz

² researcharchive.lincoln.ac.nz/dspace

OARiNZ^(w) project. As a result, they became part of that proposal and that was when they really started to think about the opportunities. Lincoln University received some modest funding from OARiNZ^(w) to support the initial setting up of the IR. The project also drew heavily on in-house talent and expertise to set up the IR.

- The IR Project team referred to a *Technical Evaluations*³ report commissioned by OARiNZ^(w) summarising an assessment of different repository software. This helped to inform their decision to use the DSpace software.
- In addition, further factors impacting the decision to use DSpace included:
 - the large number of other New Zealand universities that were using DSpace
 - the assessment conducted internally by the project team, which found that the software would be suitable for their needs.
 - being able to migrate data out of the repository in the future and being in control of their own data
 - there was also a general preference to use open source software.
- Lincoln University had been using the Australian Digital Theses (ADT) database for approximately six months before becoming involved with OARiNZ.
- Lincoln University also worked with the Ira project initially, including attending a combined meeting in Canterbury. This experience was very useful, especially as it was held at about the same time that the Lincoln IR went live.
- To formally establish the IR pilot, the University Librarian wrote a paper for the senior management group. In addition, the project team developed policies and procedures for the pilot. At the conclusion of the pilot, the University Librarian presented a further paper to the senior management group to explain the outcomes of the pilot, requesting formal establishment of the IR.
- The IR Project was included in the university planning process for 2006, and was acknowledged as a university objective. While this in itself was not difficult, it was difficult to ensure the university understood what the IR was, what it was going to offer and do.
- The benefits of the IR that were outlined during this process included preserving the university's research over time, as well as exposing and marketing the research. Providing a single interface to access research outputs of Lincoln University was also important.
- The pilot involved voluntary deposit for theses. This got limited uptake for a number of reasons, including that it was an extra thing for students to do. However, uptake was sufficient to test the processes.
- Around the same time as the pilot was underway (during 2007), the IR project started the process to support mandatory deposit of theses based on the experiences at other universities. This required a change to the regulations. As part of this change, it went to the academic board where there was very lengthy discussion. The decision was carried over two meetings as there was quite a lot of discussion about it and the project team followed up discussions with some academics individually.
- Academics expressed concern about copyright, particularly third party

³ www.oarinz.ac.nz/oarinzwiki/index.php?n=OARiNZ.TechnicalEvaluations

copyright. They felt there were issues for their students in having to get permission for third party copyright. Additional concerns were that the project was asking too much of students, and that students might not have the skills needed to convert them into pdf.

- The academic board approved mandatory deposit, starting in 2008, and the project team prepared a series of workshops with postgraduate supervisors. This gave the supervisors the opportunity to raise any questions or issues they may have had.
- Deposit of all other items is voluntary.

3. Repository Content and Usage

- The *Lincoln University Research Archive*⁴ contains 787 items (March 2009), ranging from journal articles and theses, to authored books.
- The Lincoln University Research Archive Policy states that the IR will contain full text material, including:
 - Theses
 - Serial publications: discussion papers, working papers, etc
 - Refereed research articles and contributions at the postprint stage
 - Refereed research literature at the preprint stage
 - Un-refereed research literature: conference papers, chapters in proceedings, etc
 - Research monographs and book chapters
 - Audio-visual presentations and datasets associated with research will be considered for inclusion on a case-by-case basis.
- When depositing items, basic metadata needs to be completed by the contributor, including title, author, abstract, and keywords. Since Lincoln started using DSpace, they have used the Marsden research codes for keywords. Most of the contributors attempt to use the codes, but usually it has to be modified by library staff before being approved.
- Contributors are also asked to deposit a licence. This signifies that they own the copyright and have the right to deposit the item and that grants Lincoln University a limited, non-exclusive licence to disseminate the item through the IR and to migrate the item (i.e., copy it) for preservation purposes.
- The priorities for the IR project team has been theses and the university's own serial publications. The team have occasionally put journal articles and other items, such as conference proceedings, where staff have specifically requested that they are included in the IR. There have also been a couple of occasions where staff have been going on an overseas visit to a conference or on a lecture tour and have asked for some of their items to go into the IR, allowing them to link to them.
- Library staff have retrospectively added theses to the IR, particularly where there have been requests for access, there is no spare copy, or the items are not available for purchase.
- The conversion of theses to pdf is still an issue for some students, particularly for those who are not based at the University; they might be working out in the community somewhere and don't have access to the pdf software.

⁴ researcharchive.lincoln.ac.nz/dspace

In these cases, the IR team offer to do the conversion for them. The IR team has also provided training on how to convert items to pdfs.

- There was some initial work to raise the awareness of the IR within the university. When the IR was first launched, the team visited some of the departments and showed it to them, including the divisional administrators to get them thinking of how they could use it instead of a website. This work hasn't continued as an ongoing role, but it was useful early on to raise awareness.
- The Acquisitions Librarian is managing the mandatory deposit process , including checking that the metadata is correct. The Digital Support Librarian is also available to support the deposit process (with formal support from IT technicians).

4. Successes

- New Zealand university libraries have communicated well on IR projects, and provided quite a bit of support for each other.
- There is anecdotal evidence that the IR at Lincoln University is a success. Staff are starting to rely on the IR as a place to put their research outputs. The team have found that staff want to know that they can point to the research archive to get the full text of publications.
- The IR project has strengthened relations between the library and academic staff, as well as the administrative secretarial structure in the university (often the people who are responsible for collating the research outputs for their department).
- KRIS is useful when marketing the IR to show academics and management how the IR works in the broader context of repository development in New Zealand. Academic staff have also expressed interest in KRIS as it lets them see what their colleagues at other universities are doing.

5. Issues and Challenges

- The main issue is staffing; ensuring that there are sufficient administrators with appropriate technical skills. The intention is to encourage more self-submission, but it is important to have the structures in place first. Once these are established, further work can occur to liaise with and train people on campus where necessary.
- There was a problem with support on the DSpace software, but the project team was able to draw on various experiences from other New Zealand university libraries.
- There are some ongoing technical questions - what to do when files are particularly large, there are lots of photographs, etc. It is hoped that these technical questions will reduce and the recently appointed Digital Support Librarian can manage them long term.
- There is some concern over the Marsden codes; they're very good in some areas, but they're a bit inadequate in others. It should be possible, with full text searching, to provide enough subject information without the need for a fixed subject vocabulary.

6. The Future

- The team was preparing a plan for the development of the IR; this has now been subsumed into the proposal to develop a larger digital plan for the library. This broader plan relates to all library digital services and resources.
- Work is underway to investigate how to better align with the PBRF process. Deposit in the IR does not remove the necessity to deposit evidence of research outputs for PBRF evidence portfolios and ResearchMaster with Divisional Publication Secretaries. The University Library and the Research Office are planning to work together to minimise any duplication of effort that is required to input research outputs in response to PBRF requirements and other administrative matters.
- Work will continue to convert theses retrospectively.

<http://www.oarinz.ac.nz/oarinzwiki/index.php?n=FromTheBeginning.TeresaChitty>

Page last modified 24 March 2009, at 11:52 AM ¶ Typeset by Wikipublisher

Copyright © 2005–2009 the OARiNZ^(w) wiki and its authors

Case Study: Massey University's institutional repository Massey Research Online

Institution: [Massey University](http://www.massey.ac.nz/) (<http://www.massey.ac.nz/>)

Repository: [Massey Research Online](http://muir.massey.ac.nz/) (<http://muir.massey.ac.nz/>)

Author: Tim Darlington, Digital Services Manager, Massey University Library.
T.A.Darlington@massey.ac.nz

Prepared: 23 March 2009

Summary

Massey University's institutional repository, [Massey Research Online](http://muir.massey.ac.nz/), has been in place since 2006 under different names and on different platforms. It uses DSpace open source software to make available digital versions of Masters and PhD theses completed at the University. The service is in the process of being extended to include journal articles, conference papers etc, but there are technical and staff resourcing difficulties with this.

Institutional Overview:

Massey University is New Zealand's second largest by number of enrolments. It is based on three sites in Auckland, Palmerston North and Wellington, with a large number of extramural students in addition to internal enrolments. Student EFT numbers are just under 20,000. Its main teaching strengths are in veterinary medicine, agricultural sciences, life sciences, business and finance. Its institutional repository, Massey Research Online, is housed and administered by the University's IT Services, with the Library responsible for administering the repository content.

The University's regulations relating to submission of theses were altered for Masters and PhD programmes beginning in 2007, to require submission of a digital copy of the completed thesis. There are currently around 350 documents in MRO, with a backlog of around 60 theses awaiting processing by Library staff. No digitisation projects have been undertaken. Current development effort is concentrated on broadening the repository beyond theses.

Timeline:

2005

- Massey joined the [RUBRIC](http://rubric.edu.au) project (<http://rubric.edu.au>) as international partner for a group of Australian regional universities collaborating on developing institutional repositories.
- Subscribed to Bepress' remotely hosted [Digital Commons](http://www.bepress.com/ir/) service (<http://www.bepress.com/ir/>) for 2006, as a stopgap measure to get Massey theses available online.

2006

- The first theses were loaded onto Digital Commons. Submission was voluntary via an online form.

- University regulations were altered to provide for mandatory submission of a digital version of Masters and PhD theses.

2007

- Based on outcomes of the RUBRIC project, Massey opted to switch from Digital Commons to local hosting on DSpace.
- The cutover took place on 31 December 2007 and the repository was renamed MUIR (Massey University Institutional Repository).

2008

- Discussions with the Massey Research Office re using a dark version of MUIR as a storage facility for the Research Information Management System (RIMS), for the purposes of PBRF reporting. No funding available for this in 2008.
- MUIR was renamed to Massey Research Online, as experience demonstrated the name MUIR gave Massey staff and students no indication of the purpose of the repository.

2009

- Options for matching up the repository with RIMS to be revisited.
- Additional staffing (part-time) available from April to work on further development of MRO.

Model:

Our aim is to adopt the [ARROW HERDC report's](#) Model 2: Research Management System to Institutional Repository (“...an input process where the data is captured into the research management system, with the data then flowing to the institutional repository.”), but so far there has been insufficient staffing resource available to implement and maintain that model.

Outreach, Marketing and Faculty Engagement

Initial work by faculty liaison librarians revealed that academic staff are reluctant to submit their research to the repository if:

1. It involves extra administrative overhead. Staff already must submit details of their research to the University’s Research Office for the purposes of PBRF reporting. Submitting also to the Library’s IR would be an extra step, which academic staff were unwilling to add to their workload.
2. They can’t submit the published version. In most cases, copyright bars submission of the published version. This will remain a significant obstacle to including Massey published research in an open-access repository.

Relationships with Research Management Systems and other systems

In light of the problems outlined above, the Library opted to make contact with the Research Office to discuss implementing a process model along the lines of ARROW/HERDC model 2. The intent is to have submission forms for PBRF reporting to the Research Office ask for an electronic copy of the research if

available, This electronic copy and associated metadata would be stored in a “dark” or closed-access repository for subsequent PBRF rounds, and Library staff would move content for which there was no copyright problem to the “light” or open-access repository. This is technically feasible and would resolve the problems outlined in the previous section, but so far resourcing to build and maintain it has not been available,

Technical Environment and Information Technology strategies - Platforms/Architectures

Software: DSpace version 1.4.2

Database: postgres

Server: virtual server using VMWare

O/S: RedHat Enterprise Linux 5

Customisations to DSpace:

- Usage statistics package
- University of Rochester’s personal researcher pages
- Embargo handling

Metadata schema: Qualified Dublin Core, modified to allow identification of Massey authors.

Harvesting: OAI-PMH standard metadata available for harvesting at <http://muir.massey.ac.nz/dspace - oai/request>

Harvested by:

- [Australasian Digital Theses Program \(http://adt.caul.edu.au/\)](http://adt.caul.edu.au/)
- [Kiwi Research Information Service \(http://nzresearch.org.nz/\)](http://nzresearch.org.nz/)

Sustainability

The Library has no dedicated staff for MRO. Responsibility for it is spread across various existing positions in Digital Services, Collection Services and College Liaison Services. Incoming digital theses are received and processed in Collection Services. The documents are turned into encrypted PDF files in the format required for inclusion in the Australasian Digital Theses Program, with a copyright statement added. It has been found that a significant proportion of the submitted theses require work to format them correctly for the repository, which has led to significant delays in making content live in the system. Training is planned to permit a greater number of staff to carry out this work.

The new position of Web Technologies Librarian in Digital Services will from April 2009 add resourcing to develop MRO further. Long-term sustainability of the IR is assured through mandatory submission of theses and the ability of existing staffing to handle loading of this content (once trained in how to do it). This ensures the sustainability of MRO as a repository of digital theses, but further development of it as a showcase of Massey research will be dependent on being able to integrate it into existing research reporting processes.

University Auckland New Zealand ResearchSpace@Auckland

(This case study is an updated summary of an earlier Case Study found at <http://hdl.handle.net/2292/2557>)

1. Please give us the following details • Name of Institution • Name of the repository(ies) and web address(s) • Author(s) of the case study (include contact details if possible) • Summary/abstract • Date prepared

<http://researchspace.auckland.ac.nz>

Prepared by Leonie Hayes – Research Repository Librarian

Date: May 2009

2. Institutional Overview Include information about your institution, size, focus, FTE, culture • Where the repository is managed from • Partnerships, collaborations • Name of the repository – any tips for others on choosing a name • Repository timeline • Mission statements • Regulatory environment within your institution

The University of Auckland is New Zealand's leading research-led University, the country's largest. It was established in 1883 and has 8 faculties, along with 30 interdisciplinary research clusters. The University Library is responsible for ResearchSpace one of the digital library products at the University.

ResearchSpace uses DSpace (Open Source Software) www.dspace.org , to manage the deposit and storage of Digital Doctoral theses. This is being expanded to include other research outputs like conference papers, journal articles and working papers.

Mission Statement

ResearchSpace is an open access institutional repository. Its purpose is to showcase and preserve the research outputs of members of the University of Auckland.

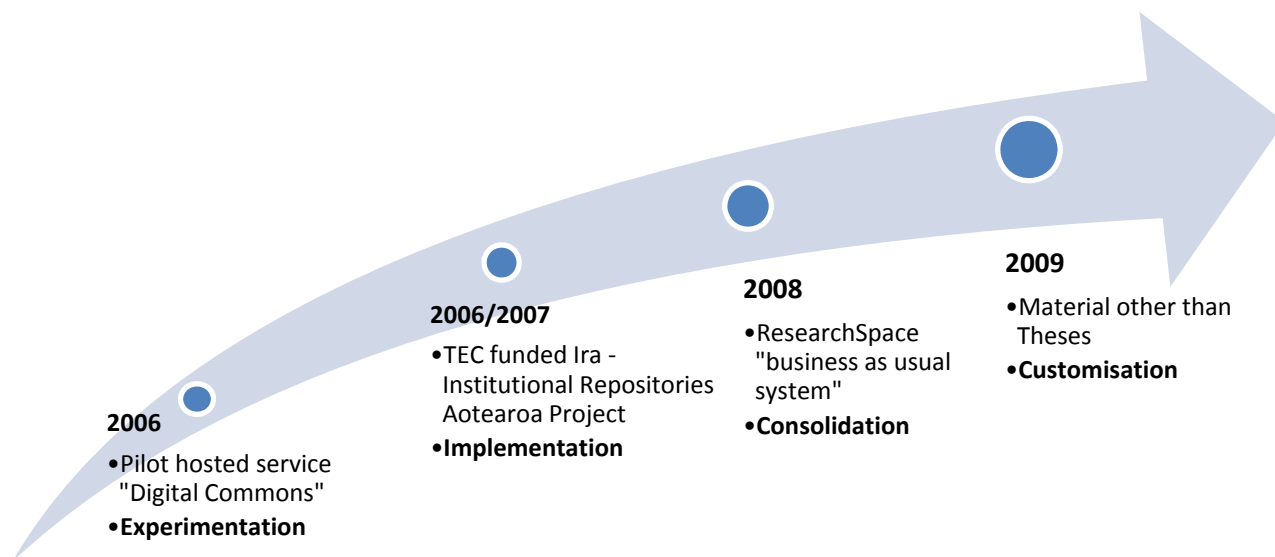


Figure 1 : Timeline

3. Comment on

- **Your institutional model based on the Arrow HERDC report**
- <http://www.arrow.edu.au/docs/files/arrow-herdc-workinggroupreport.pdf>
- **Institutional embedding and innovative Practices**

The current model for ResearchSpace is based on deposit of PhD theses, the University would like to move forward with management of other Research Publications using the HERDC style model 2.

“Major Achievements

- Partnerships with other Universities in the Institutional Repositories Aotearoa (Ira) project.
- Partnerships within the University with the Information Technology Directorate and the Graduate Board of Studies.
- Team members awarded a General Staff Excellence Award in Innovation.
- Sustainability within the existing operational budget of the University Library.
- Institutional Embedding and Business Process Analysis.
- International exposure for Doctoral Level Research undertaken at the University of Auckland. Time Line 2006 Pilot hosted service "Digital Commons" Experimentation 2006/2007 TEC funded Ira - Institutional Repositories Aotearoa Project

4. Outreach, Marketing and Faculty Engagement Give details about any of the following • Strategies and Services • Support required • Workflow • Frequently asked questions • Copyright, licencing,(include use of Creative Commons) and legal aspects • Barriers to moving forward

- Deposit of digital theses is a requirement for newly enrolled PhD students from 2007, Deposit of Masters level theses is currently being investigated. Deposit instructions and details of our use of Creative Commons is available from <http://researchspace.auckland.ac.nz/handle/2292/2>
- In partnership with the Universities copyright office a Deposit Licence to cover other materials has been developed. This is available from <http://researchspace.auckland.ac.nz/handle/2292/2857>

5. Relationships with Research Management Systems and other systems • Interoperability, data exchanges, and ongoing relationships • Data duplication and/or matching • Single sign on and authentication issues • Output to personal web pages, CV creation, Bibliographies and Social Networking tools

The relationship with the Universities Research Management systems is in development during 2009. A small collection of published journal articles and conferences papers cleared for Open Access has been made available.

A project to investigate and implement enterprise authentication using the Universities Single Sign on based on Shibboleth will be completed during 2009.

6. Research Assessment and Reporting Comment on • How your repository supports formal research assessment exercises • Other reporting exercises generally • Discuss on metrics etc outside of these formal exercises

The NZ research assessment exercise PBRF is due in 2012, the requirements for NZ Institutional repositories to contribute to this exercise have not been finalised.

7. Technical Environment and Information Technology strategies -Platforms/Architectures • Software Platform(s) • Database and Operating Systems • Unique setups • Mirrored Systems, Redundancy and backup strategies • Customisations – how heavily is your system customised? • Metadata schemas • Open URL linking • Harvesting

ResearchSpace uses DSpace 1.5.1 , a PostgreSQL database and runs on the Universities ITS Enterprise Sun E25K. The system is highly available and is mirrored at a second DR site. Details about this are available from <http://hdl.handle.net/2292/2503>

The system is lightly customised in terms of DSpace software, the major customisations have been to enable the mirroring of our service using a complex system of networking software to enable the application to have 2 internal systems but appear to the world as a single virtual location.

Dublin Core metadata schema with some additional NDLTD ETD-MS for theses has been used. An expanded metadata set to capture separate elements of journal citation data like volume, issue and pages has been used.

At this stage hard coded Open URL links to the SFX system have been implemented, in the future this will be automated.

All NZ repositories have collaborated on common metadata requirements for harvesting by KRIS - Kiwi Research Information Service <http://nzresearch.org.nz/>

8. Sustainability • Organisational Structure - Staffing and resources • Funding and external grants • Management structure – reporting lines • Rate the quality of your metadata • How much Open Access full text does your repository contain? – is it a question of resources or copyright?

The University Library has allocated 3 FTE staff members: 1. Repository Manager – Librarian; 2. Business Analyst – Technical IT specialist; 3. Repository Administrator.

ResearchSpace is an ongoing “business as usual” system and one part of the Digital Libraries suite at the University Library. At this stage 99% of content is full text, however a substantial portion of the PHD thesis collection (770) is unavailable pending consents from authors.

9. Lessons learned and Future Plans • What is next? 1 year? 2 years? Longterm • What has been easy to solve? • What has been difficult to solve? • What advice would you give others

During 2009 the XML version of DSpace - Manakin will become the User Interface for ResearchSpace. Currently a large number of technical reports published by faculty are being imported into ResearchSpace. These reports are not formally published but have been available online from faculty websites. Easy deposit of these materials using SWORD (an ATOM based published protocol for repositories) is being implemented in partnership with the Computer Science Faculty.

Advice for others, lightly customise your system if you are using Open Source software. Keep up to date with the latest versions of the software, and then you can take advantage of new and improved features. Concentrate on content if your resources are limited. Use pilots where possible to test features and functionality.

Institution: Victoria University of Wellington (New Zealand)

Repository Name: ResearchArchive@Victoria

Repository URL: <http://researcharchive.vuw.ac.nz/>

Contact: Emma Shephard-Walwyn (emma.shephard-walwyn@vuw.ac.nz)

1 Institutional Overview

Victoria University College was founded to celebrate the Diamond Jubilee of Queen Victoria in 1897, and became Victoria University of Wellington (VUW) in 1962. In 2005 the Wellington College of Education was disestablished and incorporated into the University. The University has continued to expand, and in 2006 the New Zealand School of Music was established as a joint venture between the University and Massey University, combining the musical programmes of both institutions.

The University now operates seven faculties, 27 schools and a number of Institutes and Centres across five campuses in Wellington.

VUW STATISTICS*

No. of EFTS	17,085
No. of enrolments	21,889
No. of staff	1,985
No. of campus libraries	5

*2007 figures

VUW FACULTIES

Architecture and Design
Commerce and Administration
Education
Engineering and Computer Science*
Humanities and Social Sciences
Law
Science

*Pre-2009 Computer Science was included in The Faculty of Science

2 Institutional Repository Overview

The research archive was established to provide long-term storage and access to University research outputs such as theses and academic papers. The repository enables research students and academics (past, present and future) to highlight their work wherever in the world they are, thereby increasing the presence and impact of Victoria University of Wellington research outputs both nationally and internationally.

The repository consists of twin archives:

- **ResearchArchive:** is the public repository, which contains all material for which we have full copyright clearance. It is accessed through the repository URL or the Library OPAC catalogue. Is also harvested by Google Scholar¹, Kiwi Research Information Service (KRIS)² and OALster.³
- **RestrictedArchive:** is the private repository that contains theses for which we do not have permission from the author to make their work openly accessible.

OVERVIEW OF REPOSITORY CONTENT

RESEARCH ARCHIVE	
Books	2
Conference Papers	24
Journal Articles	33
Other	14
Working Papers / Technical Reports	57
Theses – Doctorate	392
Theses – Master	260
RESTRICTED ARCHIVE	
Theses – Master & Doctorate	1218

¹ Google Scholar <http://scholar.google.co.nz/>

² Kiwi Research Information Service (KRIS) <http://nzresearch.org.nz/>

³ OALster <http://www.oalster.org/>

No other material is held in this archive. The RestrictedArchive is only available to VUW staff and students via the Library OPAC catalogue or through the Interloan system.

Material regularly moves from the RestrictedArchive to the ResearchArchive as we receive copyright clearance from authors. RSS feeds of new deposits are available from the ResearchArchive and KRIS.

The Library also has two development repositories that are used for development, testing and training. They hold no valuable information, with content able to be altered, deleted or rearranged arbitrarily.

3 Institutional Repository Development

In 2006, an LCoNZ project was established to evaluate various institutional repository software applications with a view to recommending one application to be used in the development of a multi-institutional repository using the existing LCoNZ infrastructure. Technical evaluation of a range of software applications confirmed that DSpace was one of the best solutions in the short to medium term for its ease of use and customisation ability. Although it was decided not to proceed with a multi-institutional repository, valuable information was gathered during this project that assisted VUW in the creation of our own repository.

With the Universities of Auckland and Canterbury, VUW was granted funding from the Tertiary Education Commission (TEC) for the Institutional Repositories Aotearoa (IRA) project to implement an institutional repository based on DSpace. The TEC funding enabled the Library to purchase a dedicated server for the VUW repository and to employ a DSpace system administrator/programmer, shared with the University of Auckland and the University of Canterbury.

In December 2006 an appointment was made to the newly established role of Digital Research Repository Coordinator. This role was established to develop policy and processes surrounding research submission to the repository as well as encouraging academics to submit content by raising awareness of the institutional repository within the University and acting as an advocate for its use.

ResearchArchive@Victoria went live in 2008 with early content recruitment well underway. The focus for 2008 was to provide information and training sessions for Library staff and to increase exposure and involvement with the repository. Once trained, Library staff approached academics in their allocated subject areas to continue with content recruitment, provide information about the repository and copyright advice.

In August 2008, the management of the institutional repository moved to Collection Management, with continued involvement from other Library staff. With this move, the repository became 'business as usual' and was fully integrated into Library workflows.

4 Collaboration

The LCoNZ and IRA projects highlighted how useful collaboration between institutions can be and how much we gain by sharing information and expertise. All institutions had similar needs to develop workflows, copyright expertise, quality controls, etc. Although each institution's needs varied according to internal policy and priorities, the bulk of the work in establishing individual repositories was able to be shared, and the quality enhanced through active collaboration and information sharing.

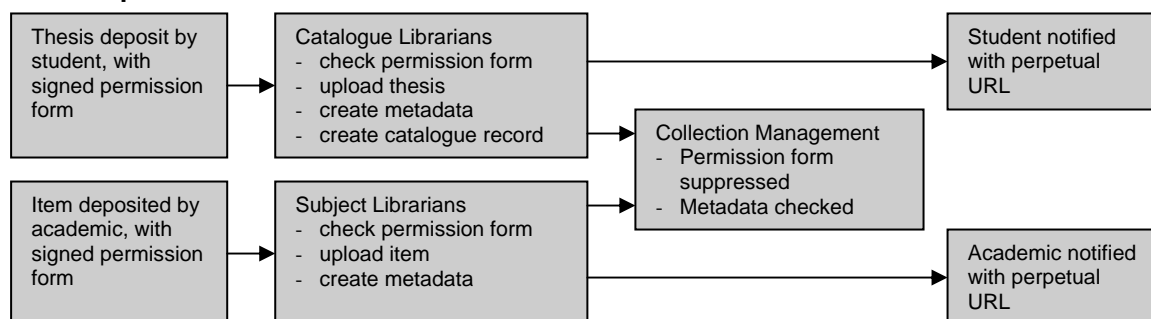
The Library continues to collaborate in the LCoNZ Institutional Research Repositories Operations Management Group to discuss and share ongoing issues, development and solutions.

The Doctorate and Masters Theses Digitisation Projects saw the collaboration of various groups within the Library, The New Zealand Electronic Text Centre (NZETC) and the University Alumni Relations Office. This collaboration continues to bring a shared understanding and to increase the skills and knowledge of all involved.

5 Institutional Repository Workflows

The repository, originally established as a project and managed by one person, has grown to become business as usual with operations being fully integrated into current workflows. There are a number of Library staff involved in the repository, each dealing with aspects of the repository that closely relate to their other responsibilities:

Item Deposit Workflow:



- **Catalogue Librarians:** are responsible for managing the student thesis deposit workflow. This includes permission forms, uploading of content to the ResearchArchive or RestrictedArchive, metadata creation and creation of a bibliographic record for the Library OPAC catalogue. Problems are escalated to Collection Management.
- **Subject Librarians:** are responsible for managing the academic deposit workflow within their allocated subject areas. This includes content recruitment, permission forms, copyright advice, uploading of content to the ResearchArchive and creation of metadata. Problems are escalated to Collection Management.
- **Lead Architect, NZETC:** is responsible for development and customisation of DSpace and managing any technical issues that may arise.
- **Collection Management Librarians:** are responsible for managing the day-to-day running of the repository and any digitisation projects. This includes determining policy, problem solving, training, copyright advice and some technical assistance.

Additionally, the VUW Institutional Repository Advisory Group provides direction on policy and developmental issues as they arise.

6 Relationship with Research Management System

The implementation of the institutional repository was driven by the Library's goal to expand its digital collections and make them openly accessible. It was a Library collection building and access initiative to support researchers and research rather than driven by the PBRF process.

Currently, the research management system used by the University, Research Master, has no connection to the institutional repository. The two systems have been built in parallel, but with few interdependencies and little interoperability. On one hand this has

meant a certain amount of duplication, on the other, it has allowed the Library to pursue its own priorities, namely the retrospective digitisation of doctorate and masters theses.

In the future we aim to build a bridge between Research Master and the ResearchArchive. By establishing a closer relationship between the two systems, we hope to increase the exposure of academics to the repository and thereby increase the number of items deposited.

7 Institutional Embedding and Faculty Engagement

The repository has achieved a significant level of institutional embedding, including the Library Statute⁴ and Library strategic planning processes. From 2008, all students completing a masters and doctorate theses must deposit both a print and electronic copy of their thesis.

University staff have been kept informed of the repository and advised of the benefits and opportunities of depositing their research. Many university staff have updated their profile pages with links to their research in the ResearchArchive. We have also made successful use of academic advocates who promote the use of the repository and can advise other staff of the benefits they have received from making their work available. We hope that in developing a relationship with the Research Management System used by VUW we will increase staff engagement.

8 Statistical Reporting

Monthly statistics are collected using Google Analytics⁵ and KRIS⁶ and circulated amongst Library staff. This information is used to update academics on the attention their work is receiving and to excite interest and 'buy-in' to the benefits of the ResearchArchive.

- **Absolute Unique Visitors:** this is the primary measure of traffic, and filters out repeat visitors (such as staff who maintain the archive) and robots.
- **Top Documents:** measures the number of visitors to the metadata page for each document. The top five visited documents for the month are included in the monthly statistics.
- **Top Searches:** measures the number of visitors coming to the website as a whole from search engines using a particular search query and the rank of the ResearchArchive on Google for that query. This is a somewhat volatile measure and is only taken as an indication.
- **Metadata Quality:** KRIS is checked monthly to ensure that our metadata remains at a high standard.

VUW STATISTICS February 2009	
Visits	7,927
Unique Visits	7,320
Page Views	21,929
New Visits	90.45%
Metadata Quality	100.00%

Outside the monthly statistics, other metrics from Google Analytics may be gathered as necessary in response to queries or to incite additional interest in the repository.

⁴ <http://policy.vuw.ac.nz/Amphora!~policy.vuw.ac.nz~POLICY~00000000021.pdf>

⁵ <http://www.google.com/analytics/>

⁶ <http://nzresearch.org.nz/index.php/>

9 Challenges

There have been several challenges that have arisen in relation to the institutional repository:

- **Pre-Publication:** increasingly students are wishing to publish their theses, and although an embargo process is in place, some believe that to deposit their thesis in the open ResearchArchive may jeopardise acceptance for publication. This situation requires delicate negotiation to resolve.
- **Other Formats:** some theses include additional material in a variety of formats, including sound recordings, video clips, spreadsheets and databases. While some formats, such as spreadsheets, can be readily converted to PDF and uploaded to the repository, other formats can be more problematic. The decision was made to add these files to the repository in their original format, provided they fall within certain size parameters. This decision does however raise some software support concerns.
- **Copyright Management:** we have made the decision not to implement different levels of creative commons licenses at this stage. We only accept items for inclusion in the repository if we have full copyright permission to do so. Subject Librarians have been trained to assist academics in clarifying author and publisher rights and to assist authors in gaining permission from their publishers.
- **Large Items:** this problem has appeared mainly during the digitisation projects. Many older theses include large fold out maps and other items that are too large to digitise with the rest of the thesis. These are digitised separately and added as additional files wherever possible.

10 Current & Future Projects

Three digitisation projects have been undertaken in conjunction with the institutional repository. The decision was made by the Library to digitise its doctorate and masters theses collections with a view to providing wider access to theses within the University and increasing the visibility of University research outputs both nationally and internationally:

- **Doctorate Theses Digitisation Project:** approximately 1,200 doctoral theses were digitised during the period March-November 2008 and uploaded to the RestrictedArchive. Part of this process included contacting authors whenever possible to gain permission to move the theses to the open access ResearchArchive.⁷
- **Large Item and Other Format Doctorate Theses Digitisation Project:** approximately 30 of the 1,200 doctorate theses contained large format items such as foldout maps and images, with another 40 having additional material on CD or CD-Rom. During the period December 2008 – February 2009, the images were scanned and added as separate files under each item record. The additional material supplied on CD or CD-Rom was uploaded and added to the item record wherever possible.
- **Masters Theses Digitisation Project:** approximately 5,000 masters theses will be digitised during a two year period starting in March 2009. This project is already underway, with the first 300 theses having been sent for digitisation.

In the future, we hope to investigate the possibility of self-deposit, and to build closer relations with Research Master, the Research Management System used by VUW.

Emma Shephard-Walwyn
Collection Management Coordinator (Acting)
30 March 2009

⁷ Further information on the Doctorate Theses Digitisation Project can be found in: Mason, I. (2008). Technical Report: Doctorate Theses Digitisation. Accessed 15 March 2009 from <http://researcharchive.vuw.ac.nz/handle/10063/415>.

Research Repository Case Studies - Educause 2009 Perth Australia

[Exit this survey](#)

1. Research Repository Case Studies - Educause Australasia Conference 2009 Perth Australia

For presentation and discussion at the "Research Repository Managers Symposium", EDUCAUSE Australasia 2009, Perth Western Australia. 3-6 May 2009 [http://www.caudit.edu.au/educauseaustralasia09/Session/Paper number 99.00](http://www.caudit.edu.au/educauseaustralasia09/Session/Paper%20number%2099.00).

Convenors

1. Leonie Hayes - University of Auckland, Auckland, New Zealand
l.hayes@auckland.ac.nz
2. Teula Morgan - Swinburne University of Technology, Melbourne Australia
tmorgan@swin.edu.au
3. Tom Ruthven - University New South Wales, Sydney, Australia
t.ruthven@unsw.edu.au

The Symposium convenors invite Repository Managers to submit a Case Study which will become part of the conference proceeding and available online.

You do not need to attend the conference to submit a case study and we encourage all Repository Managers to contribute to this exercise.

For conference attendees, there is opportunity during the Symposium for Repository Managers to discuss/share and communicate to others.

The survey gives guidelines on suggested topics to include in your case study. You can also prepare it based on the template available from:

<http://hdl.handle.net/2292/3368>

Deadline: 30 March 2009

1. Please give us the following details

- **Name of Institution**

Waikato Institute of Technology (Wintec)

- **Name of the repository(ies) and web address(s)**

Wintec Research Archive

<http://researcharchive.wintec.ac.nz/>

- **Author(s) of the case study (include contact details if possible)**

Sarah-Jane Saravani, Learning Hub Manager, Wintec, sarah-jane.saravani@wintec.ac.nz

Theresa Ball, Electronic Resources Librarian, Wintec, Theresa.ball@wintec.ac.nz

Ellis Gatchalian, Systems Librarian, Wintec, eliseo.gatchalian@wsintec.ac.nz

- **Summary/abstract**

The Wintec Research Archive was established in late 2008 to accept the scholarly and intellectual outputs of staff and students at the Waikato Institute of Technology, Hamilton, New Zealand. The archive is designed to reflect the PBRF process that operates across a number of research-active institutions within New Zealand, with a view to exporting data easily into any template or software that is used for the next PBRF round in 2012.

The repository software deployed is ePrints, we are investigating the possibility of using SWORD protocol for ease of import and export of data.

- **Date prepared**

27 March 2009

2. Institutional Overview

Include information about your institution, size, focus, FTE, culture

Wintec is one of the largest Institutes of Technology/Polytechnics ("ITPs") within New Zealand. The institute's reputation as a leading provider of high quality applied, vocational and professional qualifications in the greater Waikato region continues strongly, with a heritage spanning over more than 80 years.

Wintec's place in the community is reflected in its mission statement: "To build a stronger community through education, research and career development."

Wintec has three campuses in Hamilton. The main city site accommodates programmes from the Schools of Information Technology, Business & Administration, Health & Social Services, Communication, Education & Social Development, Science, Te Toi-a-Kiwa: Maori, Pasifika and Indigenous

Studies, English Language and Media Arts as well as most of the central administrative services.

The 57.5 hectare Avalon campus on the northern outskirts of the city was purpose built for our specialist programmes. This campus houses the Schools of Trades, Engineering & Construction, Sport & Exercise Science, and Retail & Service Industries. The third Hamilton campus, the Horticultural Educational Centre, is situated amidst the 58 hectares of the Hamilton Gardens.

In addition Wintec has two regional centres, in Te Kuiti (with the Otorohanga Trades Training Centre as a sub-campus) and Thames and an office in Beijing, China, which opened during 2004. Other programmes are delivered from various sites throughout the North Island.

Wintec enrolls over 26,000 students in two faculties - the Faculty of Business & Technology and the Faculty of Health, Arts & Social Sciences. These faculties encompass 12 schools of study and over 300 FTE academic staff. The institute offers programmes across all levels, from certificate and undergraduate to graduate and post-graduate degrees.

- **Where the repository is managed from**

Wintec Library, hosted on a local server supported by the ITS department

- **Partnerships, collaborations**

Wintec also hosts the research outputs of another Institute of Technology and has the potential to offer this service to others across the sector.

The repository has been registered with the Kiwi Research Information Service site (KRIS) run by the National Library of New Zealand, also on the mega repository sites OpenDOAR and ROAR.

- **Name of the repository – any tips for others on choosing a name**

Wintec Research Archive

Chose a name that will not date quickly, is not in use by another institution, demonstrates clearly the purpose of the repository, is neutral in expression, unambiguous and not too long.

- **Repository timeline**

Initial development using ePrints to establish a Learning Objects repository – 2005/06. Redevelopment of ePrints repository as research archive early–September 2008.

Listing repository on external sites November –December 2008.

Hosting of external clients December 2008-February 2009.

Ongoing development and expansion of repository, (deployment of additional protocols, e.g. SWORD, interactive help files e.g. MediaWiki, **collection** building software, e.g. Greenstone, portfolio software e.g. Exabis) February 2009 –

- **Mission statements**

The Wintec Research Archive will maintain and make accessible the scholarly and intellectual outputs of staff and students of the Wintec research community.

Rationale:

EPrints was selected as repository software of choice during 2006 in a TEC-funded project to establish a learning objects repository. We looked at the usual range, including Fedora, D-Space, and decided ePrints, supported by the University of Southampton, met our requirements. These were:

- **International:** Guided by conformance to international recognised standards. Has global penetration and recognition, with an identified, open-standards based, and developmental road-map. There is an active international code-contributing and testing community. In essence, the deployment would align well with international standards and leading edge technical developments.
- **Technical:** E-Prints is “open-source” based on familiar LAMP architecture (Linux, Apache, MySQL and Perl/PHP). This architecture was familiar, being the foundation model for the LMS Moodle, also deployed within Wintec, thus allowing staff who have the ability, technical skills and required experiences to seamlessly deploy a “bespoke” solution for Wintec.
- **Support:** The application came pre-packaged with extensive, indexed and structured technical and user support documentation, significantly reducing training and familiarisation costs for depositors and viewers. This is enhanced by the active international user community.
- **Financial:** The developmental road map, the vibrant community of practice, the extensive technical documentation and user help files available meant e-Prints was acknowledged as the most economic solution to meet Wintec’s requirements.
- **Maori:** The “open-code” ensures interface translations can be undertaken (e.g. Maori Language pack) ensuring institutional obligations to the Treaty of Waitangi and the local iwi are addressed.

- **Regulatory environment within your institution**

Wintec is a Microsoft environment but also supports the development of open source products, a number of which require Linux hosting.

3. Comment on

- **Your institutional model based on the Arrow HERDC report**

- <http://www.arrow.edu.au/docs/files/arrow-herdc-interimreport-june08.pdf>

The process of submission of research outputs at Wintec follows the Model 4 outlined in the ARROW HERDC interim report. The research management system is currently in consideration but will be expected to draw upon the data held within the ePrints repository and, potentially, sit as an integrated layer within an overall institutional collection management architecture.

- **Institutional embedding and innovative practices**

This is a work in progress. It is anticipated the Research Archive will sit within a larger concept of an eScholarship Centre, which will be the online central point for a range of different collections, activities and projects across the institution. At present these collections and activities are scattered or not recorded. It is also intended that ePrints will interface with Wintec's Learning Management System, Moodle, to allow seamless access for learners to the variety of resources contained within the repository.

4. Outreach, Marketing and Faculty Engagement **Give details about any of the following**

- **Strategies and Services**

The archive was officially launched at a ceremony to celebrate the completion of the first MA Nursing theses at Wintec. These were presented to the Library in hard copy and also to the repository in electronic format. This generated initial publicity for the repository.

Subsequent presentations have been held with members of Executive, including the CEO, with the Heads of School forum, with the various Schools, through the Research Office seminar series and through individual approaches to researchers.

- **Support required**

Technical requirements for the repository are supported by staff in the ITS department. Support for ePrints source code development is undertaken by Library staff, support for editing submitted entries is also undertaken by Library staff. Marketing the repository and engaging researchers is a Library activity as is planning future developments around the activities of collecting, maintaining and providing accessibility to digital resources.

- **Workflow**

After appropriate training, the research output is submitted by the creator of the output. The submitted item is held in an editing area where it is reviewed by a member of the Library staff. Upon completion of this stage, the output will either be returned to the submitter with request for additional information or, if there are no further queries, submitted to the live repository, where it appears after the overnight refresh.

Ongoing work/developments on the archive is undertaken by Library staff with expertise in electronic resources and systems requirements.

The Library staff also complete the above activities for the hosted Institute of Technology client.

- **Frequently asked questions**

What is the Wintec Research Archive?

The Wintec Research Archive is an open-access collection of research produced by The Waikato Institute of Technology (Wintec) staff and students.

Its purpose is to showcase and preserve the scholarly work generated at Wintec. A majority of the content is available as full-text to ensure the widest possible dissemination of the material. Works include journal articles, theses, conference papers, research projects, notification of performances etc.

The Wintec Research Archive is administered by the Wintec Library.

What are the advantages of contributing to the Archive?

- The Archive is searched by Google, Google Scholar, OAlster and other search engines, increasing the chances of your research being found by anyone around the world.
- Your research can be read and shared with the International community if contributions are made in full text.
- By increasing access to your research through the search and full text capabilities, your research will have an increase in the number of times it is read and cited by others. This will raise your research profile and standing within the academic community.
- All items within the archive are regularly backed-up and provide a link to your research that won't change over time.

Will my material be found via a Google search?

Yes. The Wintec Research Archive uses [Eprints](#), an [OAI](#) compliant open source platform that is searched by Google.

Who can contribute to the Archive?

Current students and staff of Wintec are able to contribute to the Archive using your computer username and password to login, see the Wintec Research Archive [Collection Policy](#) for more information.

Is contributing to the Wintec Research Archive the same as publishing?

Contributing to the Archive will enable your research to be published online and therefore give protection over your intellectual property. However, this form of archiving is seen as complementary to the formal procedure of getting published in a journal. While contributions to the archive are checked for quality, the esteem and recognition that comes from publishing in a journal is greater.

What can I contribute to the Archive?

The Wintec Research Archive represents the research and scholarly activity of Wintec, therefore all material contributed to the archive should meet acceptable standards of intellectual quality. See the Wintec Research Archive [Collection Policy](#) for more information.

Which file formats are able to be included in the Archive?

pdf is the preferred file format for contribution to the archive. This is recommended so that compatibility over time is maintained, meaning that your research will be able to be accessed in the future.

Other formats will be accepted as required by the content, e.g. jpeg, gif

If your file is not in one of these formats, please contact the [Archive Administrator](#) to discuss the feasibility of the Library converting the document for you.

All images should be optimised for the web prior to loading into ePrints.

What if I don't have an electronic copy of the document?

Because of the varied nature of the research carried out at Wintec, it may not be feasible to have an electronic copy of your research. You are still encouraged to contribute the abstract or details of your research to facilitate the PBRF process. If you own the copyright to the item, you may also be able to scan the original item for contribution.

What about copyright?

In order to legally contribute to the Wintec Research Archive, you must hold the copyright for the material. If you have published the material elsewhere you may have signed an agreement that assigns copyright to the publisher. Please see the [copyright guidelines](#) or contact the [Archive Administrator](#) for more information.

Who do I contact with questions about the Archive?

Please direct all queries about the Wintec Research Archive to the [Archive Administrators](#).

• Copyright, licensing, (include use of Creative Commons) and legal aspects

Ownership of material within the Archive is retained by the copyright holder(s). Items in the Archive can be used for the purposes of private study and research within the terms and constraints of the appropriate [Creative Commons Licence](#), or the Copyright Act. See the [Copyright Guidelines for Academic Staff and Students](#) for further information on how to use material within the Copyright Laws of New Zealand.

To the best of our knowledge, all material available through the Wintec Research Archive is with the permission of all rights holders. If you believe material within the Archive has infringed on your rights, please [contact us](#) and we will take steps to rectify the situation.

Copyright for submissions

The author/creator needs to ensure that at least one of the following copyright conditions is met before submitting it to the Archive:

- author/creator holds copyright to the material;
- written permission has been gained from the publisher; or the
- publisher allows the author to self-archive the material in an institutional repository.

A majority of academic journals allow authors to self-archive a copy of their article. In some instances you may be restricted as to which version or draft of the item you can submit to the archive. To find out the policy of your journal publisher, search the [SHERPA RoMEO](#) database.

If there are any restrictions, access to the full text version will be blocked, but the freely accessible information describing the material will still increase the visibility of your work and make the PBRF process easier for you.

PrePrint

A preprint is the early version of an item to be submitted for publication. Because it is an early draft, it may contain errors.

Postprint

A postprint is the version of an item after it has been edited and accepted for publication. Depending on the restrictions placed by the publisher, the version you can contribute might be your version prior to being formatted for publication, or the final version that appears in the journal.

• **Barriers to moving forward**

The main barriers to moving forward are resourcing – time, staffing, financial. Our plans for development are large, our allocated resources are not. We are also keen to gather more involvement from researchers across the institution but recognise the many demands on their time as well.

5. Relationships with Research Management Systems and other systems

This area is currently a work in progress. We look forward to being able to make comments on these issues in the future.

- Interoperability, data exchanges, and ongoing relationships
- Data duplication and/or matching
- Single signon and authentication issues
- Output to personal web pages, CV creation, Bibliographies and Social Networking tools

6. Research Assessment and Reporting Comment on

• **How your repository supports formal research assessment exercises**

Each year Wintec reports on the research activities undertaken by staff in the form of a research register. This register is completed at the end of the year and so, by the time of its publication, is retrospective. The research archive will relieve researchers of the necessity to gather all the past year's outputs together at the end of the year and report on them. They are actively encouraged to submit to the repository in real time.

The repository has been designed to support Performance-Based Research Fund (PBRF) requirements, with a view to encouraging researchers to enter their outputs once only and have them maintained securely throughout the period of review (currently 6 years).

7. Technical Environment and Information Technology strategies -Platforms/Architectures

- Software Platform(s)
 - Linux
 - mod_perl
 - PHP
 - XML
- Database and Operating Systems
 - MySQL 5.0.x
 - Apache 2.2.x
 - Open Suse
- Unique setups
 - Added antiword
 - Added AWstats
- Mirrored Systems, Redundancy and backup strategies
 - Backup Script created
 - Backup located on /opt/eprints/bacup/eprints
- Customisations – how heavily is your system customised?
 - CSS Modifications made for interface
 - Integration on LDAP to support Authentication
 - Perl Script modification to support specialised functions.

• Metadata schemas

abstract:AbstractFrom

accompaniment:AccompanimentFrom

book_title:Title of BookFrom

commentary:Commentary on *Core Field* completion_time:Completion TimeFrom

composition_type:Composition TypeFrom
conductors:ConductorsFrom
contact_email>Contact Email Address*Core Field*contributors:ContributorsFrom
copyright_holders:Copyright HoldersFrom
corp_creators:Corporate CreatorsFrom
creators:CreatorsFrom
data_type:Type of DataFrom
date:DateFrom
date_type:Date TypeFrom
datestamp>Date Deposited*Core Field*department:DepartmentFrom
dir:Directory*Core Field*divisions:DivisionsFrom
documents:Documents*Core Field*editors:EditorsFrom
eprint_status:Item Status*Core Field*eprintid:Item ID*Core Field*event_dates:Event DatesFrom
event_location:Event LocationFrom
event_title:Event TitleFrom event_type:Event TypeFrom
exhibitors:ExhibitorsFrom
fileinfo:Links to files*Core Field*full_text_status:Full Text StatusFrom
funders:FundersFrom
id_number:Identification NumberFrom
importid:Import Collection*Core Field*institution:InstitutionFrom
isbn:ISBNFrom
ispublished:StatusFrom
issn:ISSNFrom
item_issues:Issues*Core Field*item_issues_count:Number of issues*Core Field*keywords:Uncontrolled KeywordsFrom
lastmod>Last Modified*Core Field*latitude:Latitude*Core Field*learning_level:Learning LevelFrom
longitude:Longitude*Core Field*lyricists:LyricistsFrom
metadata_visibility:Metadata Visibility*Core Field*monograph_type:Monograph TypeFrom
note:Additional InformationFrom
num_pieces:Number of PiecesFrom
number:NumberFrom
official_url:Official URLFrom
output_media:Media of OutputFrom
pagerange:Page RangeFrom
pages:Number of PagesFrom
patent_applicant:Patent ApplicantFrom

pedagogic_type:Pedagogic TypeFrom
place_of_pub:Place of PublicationFrom
pres_type:Presentation TypeFrom
producers:ProducersFrom
projects:ProjectsFrom
publication:Journal or Publication TitleFrom
publisher:PublisherFrom
refereed:RefereedFrom
referencetext:ReferencesFrom
related_url:Related URLsFrom
relation:Relationships*Core Field*replacedby:Replaced by*Core Field*rev_number:Revision*Core Field*series:Series NameFrom
skill_areas:Skill AreasFrom
source:Source*Core Field*status_changed:Last Status Change*Core Field*subjects:SubjectsFrom
succeeds:Later Version of*Core Field*suggestions:Comments and SuggestionsFrom
sword_depositor:SWORD Depositor*Core Field*slug:SWORD Slug*Core Field*task_purpose:Task PurposeFrom
thesis_type:Thesis TypeFrom
title:TitleFrom
type:Item Type*Core Field*userid:Depositing User*Core Field*volume:VolumeFrom

8. Sustainability

• Organisational Structure - Staffing and resources

The staffing allocated to running and developing the repository and add-on products are:

Library – Library Manager, Electronic Resources Librarian, Systems Librarian, Flexible Delivery Librarian

ITS – Technical Services Team Leader

• Funding and external grants

The initial attempt at creating an ePrints repository was funded in 2005/06 through a Tertiary Education Commission grant. The resurrection of the repository in 2008 and onwards has been funded entirely through routine, operational funding. At present, most of the planned initiatives will focus on open source products, although these carry costs, not always in the range of proprietary products.

• Rate the quality of your metadata

A mix of Dublin Core standards and user supplied keywords give flexibility as well as compliance.

• How much Open Access full text does your

repository contain? – is it a question of resources or copyright?

Approximately 50% of the material is full-text with copyright compliance being the reason why more of the resources aren't available as full-text.

9. Lessons learned and Future Plans

• What is next? 1 year? 2 years? Longterm

We are looking at using the ePprints repository as the base building block for a range of products and services.

Currently we are investigating adding on the SWORD protocol to allow ease of input and export.

We are also working with MediaWiki to develop user-friendly help files.

We will be investigating close integration with Moodle, and the possibility of interfacing with Greenstone to allow development of resources from the repository.

We plan on adding an e-portfolio system to allow researchers to develop their research profiles.

More long-term plans include the addition of an open-source e-publishing platform to highlight activities around repository resources.

• What has been easy to solve?

At times there were communication issues between the library and the programmers for the Archive, this was resolved with an increased number of meetings and further clarification as to what was expected from the Archive.

• What has been difficult to solve?

Access to server space and ITS expertise has been difficult due to funding and time constraints experienced by those departments. This was beyond the control of the team working with the project and resulted in the programming for the Archive also moving to the library.

• What advice would you give others

Prepare a business plan and persuade the members of your executive of the importance to your institution of the initiative. Allow adequate resourcing to establish and maintain the repository, if you intend developing it yourself.

Work with other areas of your institution to bring in required expertise.

Encourage your staff to develop skills in the area. Be aware of what others have done or are doing in the area, build networks and be prepared to share.

10. Additional Information

Add other information that you would like to share, include links, other documentation and resources.

Eprints Web Configuration Management – Administration view

<http://www.eprints.org/software/training/configuration/EPrintsWebConfig.ppt>

Research Repository Case Study: The University of Waikato Library, Hamilton, New Zealand

Research Commons: <http://waikato.researchgateway.ac.nz/>

Prepared by Kate Nixon – Research Repository Coordinator
March 2009

Abstract

This case study reviews the background and features of the University of Waikato’s institutional research repository, Research Commons. It outlines repository development in its institutional context and with regard to its development in collaboration with the Library Consortium of New Zealand. The repository’s relationship with the University’s Research Management System and academic staff is discussed, along with its technical specifications, resourcing, challenges and future plans.

Contents

1	Institutional overview	2
1.1	Timeline.....	2
1.2	Mission Statement	3
2	Relationship with the Research Management System	3
3	Outreach, Marketing and Faculty engagement	4
3.1	Marketing.....	4
3.2	Common concerns	5
3.3	Workflow.....	5
4	Technical Environment and Information Technology strategies – Platforms/Architectures	6
4.1	Technical Specifications	6
4.2	Customisations.....	6
4.3	Metadata and OAI Harvesting.....	7
5	Sustainability	7
5.1	Staffing and Resourcing	7
5.2	Open Access	8
6	Lessons Learnt.....	9
7	Plans for 2009	9

1 Institutional overview

The University of Waikato is a young university, established in 1964. It has over 12,000 students, seven faculties and several research institutes. Waikato's research interests are particularly strong in computing and mathematical sciences, biological sciences, management and education. The University is also strongly focused on its relationship with Maori and on supporting Maori and Pacific research.

Research Commons is managed by the University of Waikato Library. It runs on DSpace open source software. The Library is part of the Library Consortium of New Zealand (LCoNZ) and repository hosting is organised through the consortium. Hosting responsibilities are contracted out to the University of Waikato ITS department and LCoNZ employs an IRR Technical Specialist, who works on site at Waikato. Currently, the University of Waikato and Auckland University of Technology are running their repositories using the LCoNZ solution.

1.1 Timeline

Early 2006	The Library begins to investigate options for an Institutional Repository
December 2006	Implementation of a repository running on Digital Commons software
April 2007	Research Commons heavily promoted to academic staff.
March 2008	Migration of the repository to DSpace software, hosted by LCoNZ.
Mid 2008	Beginning of Phase 1 of the Integrated Institutional Repository project.
October 2008	Funding to employ assistants to accelerate repository population.

In early 2006, the University of Waikato Library became interested in options for starting up an Institutional Repository service. Around the same time LCoNZ began investigating options as well. Intending to go with the LCoNZ solution once it was decided upon, Waikato nevertheless went ahead with the temporary solution of starting the repository using the Digital Commons software, then provided by Proquest.

Waikato had been using the Australasian Digital Thesis program to deposit and store theses since the beginning of 2006. Without the need to focus on theses, Research Commons has always concentrated on other university research outputs. These are principally journal articles, book chapters, conference papers, working papers and technical/commissioned reports. In April 2007 Research Commons was heavily promoted to academic staff and population of the repository began.

Throughout 2007 and early 2008 the LCoNZ libraries collaborated with Waikato ITS on the process of setting up DSpace instances and agreeing on common features and procedures. Research Commons was migrated to DSpace 1.4 at the end of March 2008. After this time Phase 1 of the high-level University of Waikato Integrated Research Repository Project began. Phase 1's purpose is to upload research objects created since 2006 to Research Commons and to provide support for the upcoming Performance Based Research Fund (PBRF)/Formative Research Exercise (FRE). As part of Phase 1, special funding was received to employ assistants to help with copyright checking and metadata creation for post-2006 research outputs. This additional resourcing has seen the content of Research Commons grow from less than 400 items in March 2008 to over 1500 items in March 2009.

1.2 Mission Statement

Research Commons is the University of Waikato's digital open access institutional repository. Its purpose is to publicise, freely disseminate, and preserve the work of the University of Waikato's scholars and ensure the research activities of staff and students are well known locally, nationally and internationally.

2 Relationship with the Research Management System

The University of Waikato is just beginning a project to integrate their Institutional Repository (run by the Library) with their Research Management System (run by the Research Office). The DSpace repository and the University of Waikato's Research Management System do not talk directly to each other. The Research Management System was developed in-house at the University to suit its needs and any solution to an integrated system will be developed in-house as well.

A detailed workflow of the two separate procedures has been developed as a first step towards integration. At present the Library and the Research Office communicate new inputs to each other by way of automated reports. These are currently checked for the requirements of each system and entered by hand to each respective database. As has already been mentioned, the repository has received funding for additional staffing to populate it with post-2006 research outputs. This provides additional support for the Formative Research Exercise being carried out in 2009.

To increase the interoperability between the two systems, the repository uses item type metadata that matches that used by the PBRF system. The two systems serve very different purposes and one of the

main differences is that the Research Office collects the published versions of all research outputs. For Research Commons, the author's postprint must be collected separately if it is required. Research Commons always links to the published version when it is available.

3 Outreach, Marketing and Faculty engagement

3.1 Marketing

A lot of energy and time was invested in promoting Research Commons early in its development.

- Library managers achieved buy-in from senior university management; although there is no mandate for depositing, it is encouraged.
- Promotional material was produced and distributed to departments. This included a poster (Figure 1) with champions from each faculty and an informative brochure. This material promoted the benefits of the repository for academic staff.
- Subject Librarians and the repository manager promoted Research Commons at departmental meetings.



Figure 1. Research Commons promotional poster

Our initial promotion of the repository to academic staff was very successful, but the time required for adequate copyright checking and inputting of the material submitted was greater than Library staff resourcing, and a backlog developed. This was exacerbated by the change from Digital Commons to DSpace in early 2008.

3.2 Common concerns

The biggest concern academic staff have is the thorny issue of copyright, particularly understanding and accepting the concept of an author's postprint. Some researchers and/or departments have been unwilling to put anything other than the publisher's PDF in the repository and it has been difficult to convince them to change their minds on this issue.

3.3 Workflow

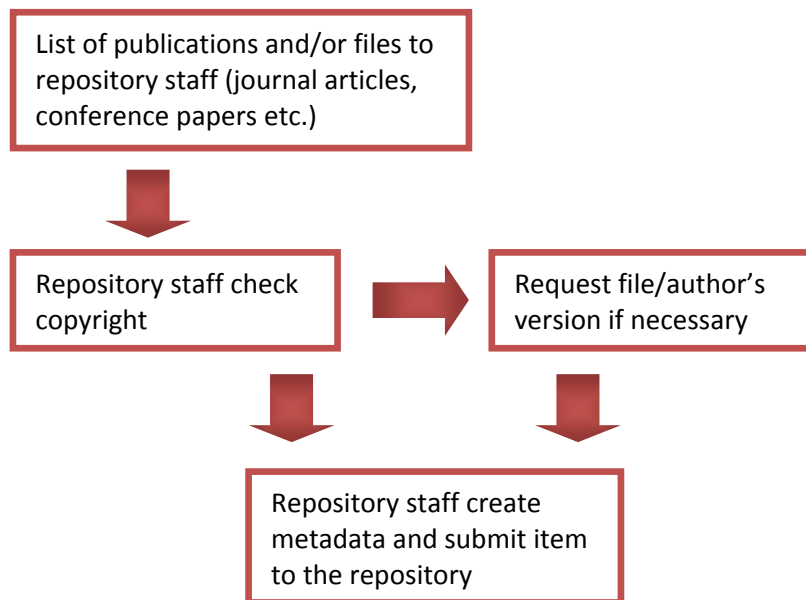


Figure 2. Research Commons Workflow

Repository workflow is fairly simple; there are no theses currently being submitted and there is no direct author self-submission into the database at this stage (although this is something expected to be looked at in the future). Knowledge of publications is obtained either through direct liaison with departments and researchers or received from the Research Office.

Copyright checking is by far the biggest and most time consuming issue. Repository staff check copyright using the SHERPA/RoMEO database (<http://www.sherpa.ac.uk/romeo/>), investigating publisher's policies on their websites or contacting publishers directly. The latter strategy is used extensively for the smaller New Zealand publishers; this can be a lengthy process but often elicits a positive response.

4 Technical Environment and Information Technology strategies – Platforms/Architectures¹

4.1 Technical Specifications

Software Platform: Research Commons is currently running on DSpace 1.5.2.

Database: PostgreSQL 8.1 (and investigating Slony-I for database replication/failover)

Operating System: Redhat Enterprise Linux 5.1

Unique Setups: The repository does not run standard Apache httpd, only Tomcat (many sites choose to run Apache with mod_jk). Tomcat service is managed by jsvc, which provides useful things like auto-restart on failure, and allows Tomcat to be run as a non-root user.

Mirrored systems, Backups: VMs make this easier, though there is no standby VM cluster in another server room. This is being investigated as a new datacentre is being built. Backups are covered in the PDF: There is an incremental each day and a full every week. Production servers are backed up by Legato, test servers by a free open source solution called Amanda. Backups are kept on disk for a short period of time then spooled to tape.

4.2 Customisations

Patches submitted for release to the DSpace community (or to-be-released):

- RSS date handling
- DOI identifiers

Patches/add-ons not submitted to Dspace, but shared with those that want them:

- Māori macron search/browse filters
- Prototype standalone statistics package, based on standard log analysis and with no patches to Dspace java source or alterations to database schema

¹ Information supplied by Kim Shepherd, LCoNZ IRR Technical Specialist. Most of this information is applicable to all the LCoNZ repositories.

- Auto-notify script for deletions (for harvesters such as the Australasian Digital Theses programme)
- Auto-notify script for archivals (this automatically reports new database items to the Waikato Research Office)

Planning/investigating:

- Custom, standalone submission interface based on SWORD or LNI (bulk ingest manager) to make submission process more flexible, and to help with any future integration with research systems
- Authority lists and functions providing interaction with dc.contributor.author objects – eg. “Contact the author”, “View this person’s research portfolio”
- Database replication with Slony-I for realtime failover of databases in the event of a DB failure, and to make database backups essentially real-time

4.3 Metadata and OAI Harvesting

Items in Research Commons are described using Qualified Dublin Core metadata fields. The repository conforms to the metadata guidelines developed by the Kiwi Research Information Service (KRIS) – <http://nzresearch.org.nz/>.

KRIS harvests Research Commons daily – details are available at <http://nzresearch.org.nz/index.php/institutions/show/67?mode=detail>

Research Commons is OAI-PMH compliant; the repository’s OAI harvest is available at <http://waikato.researchgateway.ac.nz/dspace-oai/request>.

5 Sustainability

5.1 Staffing and Resourcing

University of Waikato:

- One full time Repository Manager, employed by the Library. Reporting to senior Library management, the Repository Manager concentrates on the promotion of the repository to university staff, recruiting content, metadata creation and schemas, copyright and policy. They oversee the non-technical administration of the repository. They also liaise with the LCoNZ

group on operational matters and future developments. Currently looks after the day-to-day workflow of the repository assistants.

- Three part-time, temporary repository assistants. Specially funded by the Integrated Research Repository project, they check copyright and enter research items into the repository. Their aim is to get as many research outputs as possible from 2006 onwards into the database to bring the repository in line with the Research Office's Formative Research Assessment exercise taking place in 2009.

When the funding for the repository assistants runs out in 2009, staffing levels will return to business as usual, which is the one full time Repository Manager.

LCoNZ:

- LCoNZ has engaged Waikato ITS to set up, host, manage and customise its members DSpace institutional repositories.
- LCoNZ IRR Technical Specialist. Employed by LCoNZ, but working with Waikato ITS, the IRR Technical Specialist manages the technical aspects of the LCoNZ members' repositories. They also engage with and contribute to the world-wide DSpace community.

5.2 Open Access

Currently, 74 percent of records in Research Commons have open access full text attached. Some authors and departments have been very resistant to the idea of putting on author's postprints and have asked us to put up records linking to the full text instead. These authors/departments are very concerned with the perceived quality of the research being presented to the world. Hopefully their attitudes may change as the repository movement becomes more mature and if use of preprints and postprints becomes more accepted. However, while the PBRF requirements still specify the published version as evidence, there may not be much change.

The percentage of records with open access full text is expected to increase with the addition of the university's digital theses to the repository in 2009.

6 Lessons Learnt

- It is important to tailor the amount of material accepted for the repository to the amount of staff time available to process it. If material is not processed in a timely manner, the repository service runs the risk of looking inefficient.
- Defining the repository's mission and policies strongly from the beginning of the process makes it easier to promote to staff and to integrate with wider institutional initiatives.
- Separating the more technical aspects of running a repository from the content management has enabled the Repository Manager to focus on content recruitment and promotion.

7 Plans for 2009

1. Implement the publically viewable standalone statistics package the LCoNZ IRR Technical Specialist has been working on. It will show download statistics for items and authors, including lists of top items. This is a feature DSpace lacks and has been the feature most asked for by academic staff. It should prove to be a useful internal marketing tool for the repository.
2. Migrate theses from the Australasian Digital Thesis program into Research Commons and set up a new submittal and workflow process on DSpace.
3. Move to Phase 2 of the Integrated Institutional Repository project; serious consideration will be given to making DSpace and the Research Management System interoperable, using a yet to be determined model.
4. In 2009 we expect to increase our internal marketing and promotional activities. These were less of a focus in 2008 due to the move from Digital Commons to DSpace and the successful marketing effort of 2007 leading to a backlog of material waiting to be loaded. The project money enabling the Repository Assistants to be employed has meant this backlog has been substantially reduced.
5. Investigate using a Manakin interface, which would allow us greater control over customising Research Commons's look and navigational features.
6. Collaborate with LCoNZ Libraries on the overall direction of the group's repositories.

Research repository case studies – Educause 2009 Perth, Australia

1. General information

Name of Institution: Cape Peninsula University of Technology

Name of repository: Digital Knowledge

Authors: Debbie Becker, Michiel Moll

Summary/Abstract

CPUT libraries are currently setting up a repository. Software has been selected and a subscription to DigitalCommons signed. Collections have been identified for digitization and investigations are taking place with regard to scanning and uploading items.

Date prepared: March 2009

2. Institutional overview

Cape Peninsula University of Technology is the result of the merging of 2 Technikons (Cape Technikon and Peninsula Technikon) in the Western Cape Province of South Africa in 2004. CPUT has approximately 30 000 students (mostly fulltime) across 9 campuses and offers a range of qualifications from Diplomas to Doctoral degrees. The focus is on practical issues and includes a segment of in-service training for each qualification. We offer qualifications across 6 faculties, namely: Applied Science, Business, Education and Social Sciences, Engineering, Health and Wellness as well as Informatics & Design.

- Management

A digitization co-ordinator is responsible for liaison with DigitalCommons with regard to the setup of the repository as well as project managing the actual digitization of unique collections. The co-ordinator is also responsible for training staff and working with the Marketing team to develop a marketing plan.

- Partnerships, collaborations

In South Africa, 10 of the 23 Higher education institutions have repositories: 8 of which use D-space. Training sessions in the installation, setup and use of D-Space for a repository have been offered by the University of Pretoria. A representative from the CPUT library attended. However, it was felt that the required level of technical support needed in order to implement and maintain the product successfully, was not available on site. A decision was taken to subscribe to a hosted solution.

- Name of repository

When the decision was taken to start a repository, all library staff were invited to join a team which would be involved with the process. This group was approached to submit suggestions for a name. A

number of suggestions were received. A list was drawn up and those names which were not suitable eliminated. The complete list, including eliminated suggestions and the reasons for the elimination were submitted to Executive Management who took the final decision. Ideally, something meaningful and catchy would have been the preferred choice. However, it was felt that the word "digital" had to be included in the name. We therefore decided on Digital Knowledge. On the repository front page on our website it will read as Cape Peninsula University of Technology Libraries Digital Knowledge.

- Repository timeline

We signed the agreement with DigitalCommons in February and have already received the second draft of our website. Basic training on uploading documents to a test site has also been received. The website should be available by the end of March and staff can then be trained in uploading documents.

Collections for digitization have been identified and basic equipment such as a scanner has been purchased. Some testing of output of scanned documents has been done and it has also determined how digitization will take place, e.g. our current bound thesis and dissertations will be outsourced while certain unique collections will be scanned. Staff are currently approaching faculty to provide them with 2008 and other theses in electronic format if possible.

Testing is taking place to determine the time needed to scan one item, complete the metadata and upload the item into DigitalCommons. Once completed, the amount of time needed for a student to complete the assigned task can be determined. It is hoped that faculty will support the project by providing students from the study fields of the items being digitized to assist with the project, failing which CPUT Libraries will employ general student assistants to help with the work.

The initial aim is to have all our Theses and Dissertations, as well as 1 unique collection uploaded by the end of 2009. The second unique collection identified, will involve the scanning of 3-D objects and therefore the purchase of new equipment. This will be done in 2010.

During 2009, library and selected faculty staff will be encouraged to create their personal web pages using SelectedWorks. This part of the project will be continued in 2010 as faculty librarians market the product to faculty.

- Mission statements

Not yet developed specifically for the repository, but CPUT Libraries does have a mission statement which this will subscribe to as being part of the Library's strategic plan.

The Mission of CPUT Libraries is

- to translate/convert the academic, social and work-related information needs of the CPUT in all its facets and endeavours into innovative, effective and efficient services, systems, knowledge resources and futuristic solutions;

- To empower clients by providing them with excellent information resources, systems, services and training that add value to teaching, research, scholarship and community work, to be successful academically and as citizens;
- To research, introduce and employ information innovations, techniques and technologies into the CPUT that reflect and respond to commercial, industrial and societal needs and areas of application, and to so educate, enskill and develop CPUT Library's clients in such techniques and technologies, and in the interests being served by them; and
- To be the core of information management specialists that delivers on the wider CPUT's requirements for effective knowledge and information support.

- Regulatory environment

1. CPUT E-Thesis policy indicates that from 2008 all theses must be submitted in electronic format.
2. A library policy governing the use and extent of the repository is being developed.

3. Comment on:

- Institutional model based on the ARROW HERDC report

None

- Institutional embedding and innovative practices

None

4. Outreach, marketing and faculty engagement

- Strategies and services

A number of faculty librarians are part of the Digitization team and are already discussing the project with their faculty members. In addition the Marketing Librarian is part of the team specifically with regard to the marketing of the repository and will take responsibility for the overall marketing strategy for the repository.

- Support required

Library Executive Management is driving the project and will ensure that all staff will contribute to the process.

- Workflow

It is still to be finalised. However, it is expected that faculty librarians and faculty staff will upload documents into the repository. An identified librarian from each faculty will manage and edit submissions. Cataloguers may be requested to assist with metadata.

- FAQs

This will be developed from the list of questions asked by the Digitisation group when meeting to discuss the way forward.

- Copyright licensing

The initial collections are the property of the University and do not have copyright implications. However, a procedure and policy must be determined for future uploading. Although not yet clearly stated, faculty should ensure copyright clearance for items they upload.

- Barriers to moving forward

Currently we are only at the very beginning of the process. The main barriers currently are learning to use the equipment, training, staff to assist with the digitization and time.

5. Relationships with Research Management systems

- Interoperability, data exchanges, etc.

Currently there is no other repository on campus. There is an active e-learning site to which we could link when ready.

- Single sign-on and authentication issues

Ideally, the system should be open and accessible to any interested party. However, there will be authentication for library staff so that they can access restricted areas of the repository. In addition certain sections of the site may be restricted access for CPUT staff and students only, and there will need to be authentication for this

- Output of personal web pages

We have selected to subscribe to SelectedWorks, another BePress product which will allow for the creation of personal web pages. There is an option for data transfer between the 2 products which will allow either links from the one source to the other, or the duplication of data on the sites.

6. Research Assessment and reporting

Currently, the aim is to digitize theses and dissertations. No further decisions have been taken in this regard. Once the product is satisfactorily enabled, discussions will take place with representatives of the Research Community as to full utilization of the product.

7. Technical environment and information technology strategies

Due to the lack of capacity within our library, we have decided to opt for a hosted solution. This takes care of all technical and operational issues from database setup to archiving, including website design,

Open URL linking and harvesting. We can customized the template for metadata input to suit our needs, but have not yet completed the investigation into the data we would like to include for certain data formats.

8. Sustainability

Relevant library staff will be assigned repository-related tasks as part of their normal duties. Funding will be necessary for the continued subscription to DigitalCommons and SelectedWorks. Because the project is fairly small and is being done from the operational budget of the library and because subscribing to DigitalCommons takes care of technical issues, there should not be a problem sustaining the repository.

10. Additional information

As can be seen from the above discussion, we are still in the initial planning and prototyping phase of the repository. However, due to the hosting aspect, this phase should pass rapidly and the first public announcement of the product will take place on 26 March to the Library Senate Committee. Shortly after this it will be displayed to various stakeholders, such as Research Promotion and the Deans' Forum. The road from then on should prove both exciting and rewarding!

Please give us the following details • Name of Institution • Name of the repository(ies) and web address(s) • Author(s) of the case study (include contact details if possible) • Summary/abstract • Date prepared

*Council for Scientific and Industrial Research (CSIR) *

CSIR Research Space - <http://researchspace.csir.co.za/dspace/>
DR Martie van Deventer (0027 12 841 3278; MvanDeve@csir.co.za),
Adele van der Merwe (0027 12 841 3944; AvdMerwe@csir.co.za)
Siphethile Muswelanto (0027 12 841 2023; smuswelanto@csir.co.za)

* The CSIR is a knowledge-based Science, Engineering, and Technology (SET) organization. One of the tangible outputs generated by the organization is explicit knowledge. The creation of a CSIR Institutional Repository is a direct result of a global trend towards open access to published research results as well as a desire by the CSIR to: position the CSIR as a major generator of knowledge and contributor to national development through its knowledge outputs; make full use of a range of communication technologies available for use when presenting the CSIR to the world and disseminating research output and to harness existing infrastructure and capacity, and to provide access to CSIR knowledge outputs. CSIR Research Space contains previously published articles and papers written by CSIR researchers. Also included are selected reports which are made available by the Operating Units that generated the documents

* 24 March 2009

2. Institutional Overview Include information about your institution, size, focus, FTE, culture • Where the repository is managed from • Partnerships, collaborations • Name of the repository – any tips for others on choosing a name • Repository timeline • Mission statements • Regulatory environment within your institution

* The CSIR is one of the leading scientific and technology research, development and implementation organisations in Africa. Constituted by an Act of Parliament in 1945 as a science council, the CSIR undertakes directed and multidisciplinary research, technological innovation as well as industrial and scientific development to improve the quality of life of the country's people. The CSIR is committed to supporting innovation in South Africa to improve national competitiveness in the global economy. Science and technology services and solutions are provided in support of various stakeholders, and opportunities are identified where new technologies can be further developed and exploited in the private and public sectors for commercial and social benefit. The CSIR's shareholder is the South African Parliament, held in proxy by the Minister of Science and Technology * The Institutional Repository is managed from the CSIR Main Campus in Pretoria (South Africa) * Collaborations between the CSIR and the University of Pretoria * The Name of the repository is CSIR Research Space. An organisation should go with a name that defines the organisation's focus area, and structure. Regulatory environments within your institution- Document management systems

3. Comment on • Your institutional model based on the Arrow HERDC report • <http://www.arrow.edu.au/docs/files/arrow-herdc-interimreport-june08.pdf> • Institutional embedding and innovative practices

Yes the model is based on the Arrow HERDC, the way collections and publications

4. Outreach, Marketing and Faculty Engagement Give details about any of the following • Strategies and Services • Support required • Workflow • Frequently asked questions • Copyright, licencing,(include use of Creative Commons) and legal aspects • Barriers to moving forward

The Institutional Repository operates as a subset of an internal database called the Technical Outputs Database (TOdB). Documents in the Repository are generated via an oracle workflow process from the TOdB. This is a function that is supported by the CSIR ICT group/team. * Barriers of moving forward currently is the lack of sufficient human resources and expertise.

5. Relationships with Research Management Systems and other systems • Interoperability, data exchanges, and ongoing relationships • Data duplication and/or matching • Single signon and authentication issues • Output to personal web pages, CV creation, Bibliographies and Social Networking tools

* The Repository is a subset of the internal document database (TOdB) *The process of generating publications from TOdB is a "copy-cat"/copy and paste/duplication process because publications are copied from TodB and paste to the repository. It is however compulsory for all CSIR Researchers to submit all research outputs into the CSIR TOdB. After following appropriate routes for example checking for copyright issues, publications are then added on the repository

6. Research Assessment and Reporting Comment on • How your repository supports formal research assessment exercises • Other reporting exercises generally • Discus on metrics etc outside of these formal exercises

* Reporting exerciese: The institutional repository has a built in statstical measure as well as AWSTATS that operated independently and is monitored by the CSIR ICT group *Metrics: Alerts that allow for spot checking

7. Technical Environment and Information Technology strategies -Platforms/Architectures • Software Platform(s) • Database and Operating Systems • Unique setups • Mirrored Systems, Redundancy and backup strategies • Customisations – how heavily is your system customised? • Metadata schemas • Open URL linking • Harvesting

*DSpace * Metadata Schemas: Dublin Core * Harvesting: Google, OAISTER, Scopus, etc..

8. Sustainability • Organisational Structure - Staffing and resources • Funding and external grants • Management structure – reporting lines • Rate the quality of your metadata •

How much Open Access full text does your repository contain? – is it a question of resources or copyright?

* Manager-to-Practitioner-to-Interns * The project is centrally/organisationally funded * Quality of metadata: Feedback is high from users * Open access full text: To date the repository has over 2300 open access full text files. All the publications are copyright cleared before they are added into the repository. The Open access database institutional repository (IR) was formed after the organisational drive towards migrating to an open source environment thus the birth of open access projects for example the development of the IR

9. Lessons learned and Future Plans • What is next? 1 year? 2 years? Longterm • What has been easy to solve? • What has been difficult to solve? • What advice would you give others

*The next phase will be to add images, videos, audio files, datasets, source codes into the CSIR IR. *At the time when the CSIR repository was launched we did experience some hiccups, what could be referred to as teething problems, for example there was a period when there was a problem with the generation statistics * Advice to other organisations/institutions: Learn from others, ask for advice from those who have walked the path and look at what others are doing, don't try to re-invent the wheel

10. Additional Information Add other information that you would like to share, include links, other documentation and resources.

Adele van der Merwe had her thesis based on the project, see the link to view her thesis

<http://hdl.handle.net/10204/2504>