

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-admentment-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/>.