

Title: Connecting our community: providing the community's digital needs

Author: Peter Green, Curtin University of Technology, Perth

Introduction

During the last ten years there have been many changes in the technological environment, particularly changes to electronic communications, which have been driven by the growth of the internet and the pervasion of the web in particular. These developments have changed client's expectations and libraries have changed how they service the information needs of their communities. In many cases libraries have been proactive in anticipating some of the changes, in other cases inventive in responding to the changes. This paper will reflect on a couple of those developments, in which the author has been involved, where access to traditional resources has been enhanced by new approaches.

From dusty to digital

When the WA Institute of Technology became Curtin University of Technology in 1987 one of the implications of the change was a focus on higher degree programs such as PhD and Master by Research. The tangible output of such research is typically a thesis and the Library at Curtin, as is common practise, became the custodian of a bound copy of each approved thesis. As is also typical the copies were kept in a closed access collection, being made available for use within the library on request or copied and supplied via Inter-Library Loan for research purposes. While this is a very acceptable approach to collecting and providing access in the traditional world, it doesn't meet the current expectations of the digital age.

Innovative work at Virginia Polytechnic in the late 1990's to provide online access to completed theses prompted the establishment of an Australian project to move these valuable, but difficult to obtain, research outputs into a more open and accessible environment. Thus the Australian Digital Theses project (ADT) began with funding by the Australian Research Council. The project's rationale was to expose this particular research to a wider and global audience through exploiting the power of the web and associated communications protocols. Exposure to the digital content was envisaged as well as to the metadata. Seven Australian university libraries under the leadership of The University of New South Wales Library participated in this project between 1998 and 2001. Curtin was one of the partners.

There were several factors that made this project difficult and a few features that proved visionary.

The funding of an Australian wide project was motivated by the desire to demonstrate that theses could be re-presented in a digital format and that if a tenable solution emerged then all Australian universities could benefit from the model. The model adopted was that of a sort of distributed database. The local university would be totally responsible for their theses and the creation of a local digital theses collection. To provide incentive for exposing their theses electronically a central metadata repository would be established to gather the metadata about the thesis, held locally, and thus provide the world wide exposure that the research deserved.

The distributed model was one that librarians could understand; after all they have been sharing metadata in the form of marc records for many years. The leap was not that big, for libraries. Indeed establishing the central metadata repository and gathering metadata, based on a lightweight, common standard, would prove the easier nut to crack. The more challenging task, and one which is still being worked on across Australia and New Zealand, is to establish the local digital theses collection.

Curtin is an innovative university and a young one. Even so changing long established processes, to allow for a local digital thesis repository, was not simple, quick or easy. Two parallel methods were applied to building the digital thesis collection. The first was backward looking and involved retrospectively seeking permission, from authors of theses already held, to scan and make available their works. The second method was to solicit the thesis from the author at the point of completion (after examination and approval) and convert the thesis from its native digital format (usually Microsoft Word or similar) into a digital format (usually PDF). These two streams progressed over a number of years, successfully building a local collection. Both methods have difficulties. Authors from previous years can be elusive to contact, though usually willing to give permission. Authors who have just finished their thesis are not usually well placed to do more work. While they understand the benefits of having their thesis exposed the deposit of an electronic copy can be too much effort after many years of hard work. To address the latter problem, which is not unique to Curtin, the university is moving to streamline the process and make the deposit mandatory (allowing for the inevitable exemptions for reasons of intellectual property etc) a carrot and stick approach that should see the library receive electronic copies of most completed theses.

The difficulties mostly revolved around the need to muster the local university community; research students, supervisors and administrators to establish the local collection of digital theses. Other challenges involving the appropriate use of copyright material and handling non-traditional theses are a work in progress. The visionary aspects were the distributed holdings of theses with a central gathering of metadata thus more effectively exposing the research to the world.

The ADT project went very well. In 2002, the Council of Australian University Librarians (CAUL) agreed to transform the project into a sustainable program. All 38 Australian university libraries are now members of the ADT Program with 25 members actively contributing digital theses. In November 2004, over 2800 digital theses were accessible through the central metadata repository and growing at the rate of 75 theses per month. At Curtin the digital thesis collection has grown steadily to around 500, around a quarter of all theses held at Curtin.

In 2004 a new project was funded by the Federal Government's Backing Australia's Ability program, again led by UNSW, to expand the central metadata repository. The project is called *Australian Digital Theses Program – Expansion and Redevelopment*. This project is addressing a couple of issues that have emerged. The central metadata repository, though growing steadily and providing access to digital content, doesn't provide a one stop shop to ALL Australian research theses. There are some 100,000+ research theses held in Australian libraries and discovering them isn't as easy as the research community would like. The redevelopment project will enhance the software to allow records from the National Bibliographic Database to be uploaded on a regular basis, providing a single metadata database for all theses, while still allowing easy access to the digital content that is being provided and is steadily growing. Another enhancement is to allow metadata to be gathered and exposed using a new protocol that has emerged in recent years; the OAI-PMH. More will be said on this topic later in the paper. Other enhancements would increase the functionality of the expanded central metadata repository. This latest project is due for completion in early 2005.

From closed to open

While the ADT project was gathering momentum a movement was also brewing that would have an impact on the work undertaken over a number of years at Curtin and elsewhere on making research theses more available. The movement may be called the Open Archive Initiative. It's impact and its breadth is too vast to easily summarise here, but the movement led to a specific activity at Curtin which will be expanded upon.

In a nutshell, and all great and developing ideas can be summarised if your nutshell is superficial enough, the Open Archive movement is about opening up access to the research of the world

from behind the access barriers that have grown up over time. In particular, the core research that wants to be open is the high quality; peer reviewed output from the world's researchers. The motivation for researchers may be two fold. They may benefit from the greater exposure of their work to other researchers and their own research may benefit from increased access to the research output of others. These things are well and good, but where does a university library come into the equation?

The earliest open archive's were subject based, the most famous being arXive (an e-print service in the fields of physics, mathematics, non-linear science, computer science, and quantitative biology) which has been active since the early 1990's. However the spectacular success of arXive has not been matched generally in other subject disciplines and a complementary type of open archive has developed which is becoming known as an Institutional Repository. These open archives are not driven by the community of researchers in their discipline but by the institutions that employ researchers, and in particular their libraries. Why?

There are various motivations for libraries to get involved in open archive development. One of the motivators for libraries is that they are generally the providers of access to the peer-reviewed literature and invest considerable budget and staff time in negotiating and delivering access to this literature. In recent times the cost and difficulty of doing this in the online environment has been well documented, sometimes referred to as the "serials crisis". However a deeper motivation may be the long term perspective that librarians bring to collecting, managing and providing access to material. The housing of this high quality material in an open archive that provides a full text, accessible version of papers that can be cited now and indefinitely into the future is very appealing. In the fullness of time these open archives may be seen to be core business of university libraries that will collect, preserve and offer access to material of value well beyond the life span of the researchers. It is this long term perspective, for which researchers don't appear to have a natural affinity (being firstly concerned with their own work, and the progression of their research and careers), but which is bread and butter for librarians, that may underpin the energy librarians have devoted to this area. In fact both librarians and researchers will probably find common ground in the improved access to research and librarians are already immersed in a resource sharing environment and have been for a long time.

Having described in a nutshell the Open Archive movement, the possible motivations of its key players and the two main streams of activity (not having mentioned open access and the changes in journal publishing) a few words about how this is being played out at Curtin.

During 2003 the Library at Curtin embarked on a local implementation of an institutional repository. As with the work on digital theses this involved mustering the local university community to form a common view point (almost impossible in a university environment) and gaining endorsement from university authorities. The technical underpinning also required some time and effort but was achieved in 2003 without fuss. The Library gained approval to conduct a pilot repository at the end of 2003, with guidance from a reference group of researchers, and at the end of 2004 had the established repository endorsed as an ongoing activity.

During the pilot phase the repository was quietly populated with suitable high quality research, garnered from researchers at Curtin and has grown to 200+ papers. It is expected that this will continue to build, in the same steady way as the digital theses collection has grown, and become over time a substantial scholarly resource.

Collecting the research papers is only one aspect of the repository, and while that would be valuable in its own right, exposing the research is a prime motivator. This is where the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) comes into play. This protocol provides a standard method for sharing metadata and in particular for the metadata from different sites to be gathered into a central metadata repository where discovery services can be deployed. This simple protocol has become ubiquitous in recent years and the exposure and gathering of metadata using it is a rapidly growing industry.

To illustrate the potential of this distributed model the reader should first look at the repository at Curtin <http://espace.lis.curtin.edu.au/> and browse the locally loaded metadata and full text. The reader should then look at the demonstrator Australian discovery service, hosted by the National Library under the ARROW project banner, at <http://arrow-test.nla.gov.au/>. This service harvests metadata from eight current Australian sites holding scholarly materials, using the OAI-PMH, and provides search and alerting services for the combined metadata set, linking back to the local sites for the digital objects. In time this may well grow to become a single place to search for quality Australian open archived research and illustrates the possibilities.

Opening the door

Inevitable the two activities that have been described in this paper, digital theses and open archives, will come together. This is already happening. Enhancements to the central ADT metadata repository will include the ability to selectively harvest theses from institutional repositories that hold a variety of content. The ARROW project is building new local institutional repository software that will cater for theses and ePrints and other types of digital object. It is likely that Curtin will have one database that hosts these contents and more in the near future. However all of these activities are slow moving and a number of years will pass before we begin to see these developments settle into a stable formation. The world is still a fast changing place.

It is also inevitable that these activities are but the tip of the iceberg. The opportunity provided by rapidly developing electronic communications will lead people to see what has been done for theses and high quality research and apply these principles to other activities. It is likely that these developments are already happening but haven't gathered momentum, yet.

Looking ever forward

In the context of the university community these developments add value to the advancement of knowledge through research, an activity at the heart of a university's purpose. The objectives of providing access to materials are also at the heart of any library's purpose. Indeed the technologies and methods may find their own unique expression but the underlying principles are emerging as a consistent base on which to build. Technical solutions are usually the easier nut to crack. Finding motivations that exist in the key players and mustering the resources of a community can be challenging. Sometimes we go where we are led by our clients, other times we lead them to places they didn't know existed. In this digital age we have the opportunity to do both.

References

- ADT-ARIIC Project - Australian Digital Theses Program. (n.d.). Retrieved December 9, 2004, from <http://adt.caul.edu.au/adtariic.html>
- arXiv.org e-Print archive. (n.d.) Retrieved December 9, 2004, from <http://arxiv.org/>
- Australian Digital Theses Program website. (n.d.) Retrieved December 9, 2004, from <http://adt.caul.edu.au>
- Australian Research Repositories Online to the World. (n.d.) Retrieved December 9, 2004, from <http://arrow.edu.au/>
- Backing Australia's Ability - Building Our Future Through Science and Innovation. (n.d.). Retrieved December 9, 2004, from <http://backingaus.innovation.gov.au/default2004.htm>
- Digital Theses. (n.d.) Retrieved December 9, 2004, from <http://library.curtin.edu.au/theses/>

Genoni, P. and Cowan, R. (2003) Bibliographic control of Australian higher degree theses: the future role of the Australian digital theses program. *Australian Academic & Research Libraries*, 34(2), 81-91. Retrieved December 9, 2004, from <http://espace.lis.curtin.edu.au/archive/00000007/>

Open Archives Initiative. (n.d.). Retrieved December 9, 2004, from <http://www.openarchives.org/>

Open Archives Initiative Protocol for Metadata Harvesting. (n.d.). Retrieved December 9, 2004, from <http://www.openarchives.org/OAI/openarchivesprotocol.html>

Steele, C. (2002). E-Prints The Future of Scholarly Communication. *inCite*, 23(10), 6-7. Retrieved December 9, 2004, from <http://www.alia.org.au/publishing/incite/2002/10/eprints.html>