

**AUSTRALIAN DIGITAL THESES PROGRAM
EXPANSION AND REDEVELOPMENT**

**A proposal for funding by the DEST Research Information Infrastructure
Framework for Australian Higher Education**

Final version, August 7, 2003

AUSTRALIAN DIGITAL THESES PROGRAM – EXPANSION AND REDEVELOPMENT

DEST FUNDING PROPOSAL

Lead Institution

University of New South Wales

Partner Institutions

Curtin University of Technology; University of Melbourne; University of Queensland; University of Technology, Sydney; ProQuest Information and Learning

Team Leader

Andrew Wells, University Librarian, University of New South Wales.
Telephone: +61 2 9385 2662. Facsimile: +61 2 9385 8002. Email: a.wells@unsw.edu.au

Project Manager

Tony Cargnelutti, Manager, Online Services Department, University of New South Wales. Telephone: +61 2 9385 1531. Facsimile: +61 2 9385 8002. Email: t.cargnelutti@unsw.edu.au

Summary:

This project will redevelop the existing central metadata repository of the Australian Digital Theses Program (ADT) [<http://adt.caul.edu.au/>] to increase its coverage and utility to the national and international research community. The repository's content will expand to include metadata about all Australian higher degree theses, whether in digital form or not. This goal responds to a clearly stated desire of the Australian research community, who report difficulty in finding information about theses. The project will identify and test software solutions to support compliance with the Open Archives Initiative's Protocol for Metadata Harvesting (OAI-PMH). The partners envisage that DEST will receive other project proposals addressing common solutions for local site management of digital theses as part of the institutional repositories demonstrator. The redevelopment of the ADT metadata repository is an essential adjunct to institutional initiatives by providing efficient search services and metadata support services. The partnership with ProQuest Information and Learning provides

opportunities for retrospective digitisation of Australian higher degree theses, enabling more rapid growth of a critical mass of online content.

Amount of funding sought:

2004: \$540,000

1 Background and context

The Australian Digital Theses Program is the outcome of an ARC grant in 1997. The Program's forerunner, the Australian Digital Theses Project was led by the University of New South Wales Library with the following partner institutions: Australian National University; Curtin University of Technology; Griffith University, University of Melbourne, University of Queensland and the University of Sydney. This project developed and implemented access to distributed databases of digital theses. ADT is a distributed system in which local institutions deposit digital theses on local servers, thereby automatically generating Dublin Core metadata, which is harvested into a searchable metadata repository for all sites. The ADT metadata repository is truly open access. From the outset, ADT was designed to be flexible and scaleable: local institutions have been able to integrate it within their existing processes and IT infrastructure, including local authentication systems. The ADT program pre-dates the Open Archives Initiative but was designed along the same distributed, open access model.

In 2002, the Council of Australian University Librarians (CAUL) endorsed ADT becoming a formal program and agreed to a business plan for its development. All CAUL members are now members of ADT and pay a small levy (\$1700) to ensure maintenance of the current operation. The revenues (approximately \$55,000) are received by UNSW Library on behalf of CAUL. These funds are used for maintenance of the metadata repository, training and assistance to members. The Program has a policy committee overseeing its operation. The Committee has representatives from CAUL, the National Library of Australia, the publishing industry (ProQuest Information and Learning), the Council of Australian Postgraduate Associations and the Deans and Directors of Graduate Schools. A technical committee reports to the Policy Committee.

The ADT Program is growing steadily. By July 2003, 21 members of CAUL were actively engaged and 1500 theses were available online. The Program has provided a vehicle for innovation in scholarly communication and has dramatically improved access to research which has been previously difficult to identify and access. The Program has had an international focus since its inception, and has played an active role in the international Networked Digital Library of Theses and Dissertations (NTLTD). The vision of the NTLTD is to provide open access to a networked international library of higher degree theses. The NTLTD has 185 members, including universities and international organisations such as Unesco and OCLC. Australia has made a contribution to the NTLTD and UNSW will host the international Electronic Theses and Dissertations Symposium in 2005.

Theses will always be a major component of an institution's digital repository. A very small proportion of theses are now published as scholarly monographs. Higher degree theses represent significant research in Australian universities. Higher degree candidates contributing to the Australian Digital Theses Program

are gaining direct experience with creating scholarship in digital form and disseminating their research through open access.

The ADT Program is developing a centralised 'one-stop shop' for digital theses through its centralised metadata repository at UNSW. However, this still represents a small proportion of all Australian higher degree theses. This project proposal recommends building this metadata repository to become a comprehensive record for all higher degree theses. Records for theses are currently held in the National Library's National Bibliographic Database accessible via the Kinetica service, and also in local catalogues and printed works. A recent survey by the National Library revealed fairly high rates of coverage of higher degree theses on the National Bibliographic Database. However, the theses are not easy to find. This proposal recommends ways to use this metadata to improve visibility and awareness of Australian higher degree theses and provide the ADT Program with a critical mass of data and a firmer platform for growth.

2 Objectives

The objectives are to:

- Build on existing awareness and use of Australian theses through efficient search services
- Achieve compatibility with the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)
- Provide expertise and advice on the management of digital theses in local institutional repositories
- Bulk retrospective digitisation of Australian higher degree theses.

These objectives cover aspects of both demonstrator projects. The provision of effective search services and metadata support addresses parts of the "Common technical services" demonstrator. Local management of digital theses and its relationship with the ADT Program addresses issues in the "Implementation and support for institutional repositories" demonstrator.

2.1 Build on existing awareness and use of Australian theses

The research community reports difficulties in discovering Australian theses: this objective addresses the *discovery* challenge through bringing metadata about all Australian higher degree theses together. The metadata currently exists in various databases, but it is not easy to search. Researchers do not always know where to start searching. It is extremely difficult to search by discipline.

The *discovery* challenge can be met by putting all existing metadata together in one repository. Stakeholders such as the Deans and Directors of Graduate Schools and the Council of Australian Postgraduate Associations have identified

the need for a specialist repository for such core research outputs as higher degree theses. The ADT repository has developed an identity and gets high usage from national and international searchers. Using this existing repository builds on the success of the ADT Program. This improvement in access will enable greater visibility and usage of Australian theses.

2.2 Achieve compatibility with OAI-PMH

Currently the ADT uses both qualified and unqualified Dublin Core metadata elements to populate its central repository. The proposed extension to the ADT includes building a simple OAI Metadata Harvesting Protocol interface to the central ADT metadata repository. Improved international visibility of Australian higher degrees will be enabled when ADT metadata is incorporated smoothly into other metadata services.

2.3 Providing expertise to management of digital theses in institutional repositories

The project partners have developed considerable understanding of the processes involved in the online submission of digital theses. This expertise will be made available to other successful bids under the institutional repository demonstrator category. The management of online theses needs to be embedded in the local academic practices and submission processes of each institution.

This project proposes testing ProQuest's online submission service for digital theses. This will enable comparison and evaluation of local software solutions with an outsourced service provided by ProQuest. Evaluation criteria in the comparison would cover issues as cost-effectiveness, access to digital theses, and metadata quality and use.

2.4 Retrospective digitisation

Increasing the content accessible through ADT can be accelerated through targeted retrospective conversion of theses. Some CAUL members have active programs to do this (eg, Curtin University of Technology). ProQuest Learning and Information also provides a retrospective conversion service for theses. The key objective here is to build a critical mass of online content accessible through the ADT metadata repository. This could give ADT the ability to significantly increase its full text offering in a timely and cost effective manner.

3 Project activities

Activities are:

- Software redevelopment of the metadata repository

- Adding metadata for theses not currently represented in the ADT metadata repository
- Enabling relationships with other national and international search services for theses
- Developing retrospective content and awareness of Australian higher degree theses accessed through the metadata repository
- Assist in the development of common local site software to handle digital theses and compare local site software with an outsourced commercial service solutions

3.1 Software redevelopment of the metadata repository

The metadata repository needs the following enhancements:

- Developing the ADT metadata repository to become an OAI-PMH compliant harvester of metadata (ie, Service Provider) as well as a source of metadata (ie, Data Provider) for other national and international metadata repositories and resource discovery mechanisms (see Appendix 3)
- Data handling tools to convert incoming formats (MARC) to Dublin Core
- Data quality and management tools (de-duplication, authority control)
- Ability to add new content fields – eg, inclusion of abstracts in addition to the bibliographic record when full text not available
- Improved search interface. Users should be able to distinguish between records linked to digital theses and those not; searching by discipline; searching for records with abstracts etc
- Improved management and reporting tools
- Support for interoperability between repositories by use of common search and retrieve protocols, eg Z39.50, SRW (Search/Retrieve over the Web)

The project will consider options for the redevelopment of the metadata repository:

- Identifying commercially available solutions which are non-proprietary
- Develop new software using open source tools

Functional specifications will be developed and tested against both options. A prototype would be developed and tested. Full implementation would follow.

3.2 Adding metadata for theses not currently represented in the ADT metadata repository

The chief source of data is the National Bibliographic Database. Around 200,000 MARC records would need to be extracted from the NBD, converted to Dublin Core and added to the metadata repository. Existing crosswalk conversion programs will be used to facilitate mapping between data formats. Other data sources include local catalogues, existing institutional repositories and ProQuest Information and Learning.

The project would involve extraction of data, testing conversion and loading the data. The metadata repository would need to be appropriately sized to cater for satisfactory performance and anticipated growth.

The inclusion of abstracts also adds value to the metadata repository. Locating abstracts in print or digital form and including them in records would also form part of this project. ProQuest is another source of abstracts and the project will investigate capturing this data and adding it to the ADT metadata repository.

3.3 Investigating relationships with other national and international search services for theses

This section of the project will address ongoing relationships for metadata exchange with local institutional repositories, national search services delivered by the National Library and international search services such as the Union Catalogue of Electronic Theses and Dissertations being developed in the USA.

3.4 Developing retrospective content accessed through the metadata repository

This section of the project will cover:

- Methods to identify key theses and disciplines warranting retrospective digitisation. This would be done through inspection of interlibrary loan requests and consultation with the research community
- Establishing pilot projects with ProQuest Information and Learning to test its retrospective digitisation service
- Comparing the cost effectiveness of using the ProQuest digitisation service with in-house digitisation
- Negotiating with ProQuest Information and Learning to acquire or license their existing Australian higher degree theses in digital form for addition to the ADT Program

3.5 Assist in the development of common local site software to handle digital theses

This section of the project envisages interactions with other project proposals submitted for both demonstrators (institutional repositories and common technical services).

The project partners will provide advice and expertise to institutional repository software projects that include digital theses submission and management. The ProQuest digital thesis online submission system will also be evaluated.

The project can work with authentication and access management proposals submitted and funded under the Common technical services demonstrator. The possibility of testing access management via AARLIN could also be investigated.

4. Outcomes

- Significantly expand Australian research information infrastructure
- Significantly enhance existing technical research information infrastructure
- Expose Australian research to national and international audiences
- Embed digital theses in institutional repository developments
- Develop partnerships with industry
- Enable universities to build digital repositories seeded by content developed in ADT
- Significantly contribute to open access at national and international levels.

5. Usefulness for higher education institutions and community at large

- Submission and management of higher degree theses incorporated with institutional repository and digital publishing systems
- Comprehensive metadata about Australian theses delivers efficient access to and visibility of research to higher education
- Open access to research

6. Dissemination

- Milestone reports through Council of Australian Postgraduate Associations (CAPA), Deans and Directors of Graduate Schools (DDOGS), Deputy Vice-Chancellors Research (DVC(R) group), Council of Australian University Librarians (CAUL), National Library of Australia (NLA)
- International announcements through lists, NTLTD
- ADT Policy Group to contribute

7. Implementation timetable

All activities in 2004 calendar year

7.1 Software redevelopment of the metadata repository

Responsibility: Project Manager and UNSW/Curtin (support from ADT Technical Committee)

Activity	Date
Develop specifications for new metadata repository	March
Issue RFP, evaluate responses, make decision to use or develop software	May

Install and test (or write and test)	October
--------------------------------------	---------

7.2 Adding metadata for theses not currently represented in the metadata repository

Responsibility: Project Manager and all partners

Activity	Date
Identify and evaluate sources of metadata; advise on data conversion tools required and server requirements	February
Negotiate extraction of data	June
Test sample data	October
Install software and hardware	November
Load data	December

7.3 Investigating relationships with other national and international search services for theses

Responsibility: Project Manager and UNSW/UTS

Activity	Date
Agreements with key services (eg National Library of Australia)	June
Test sample data	November
Test gathering by other OAI service providers	November

7.4 Developing retrospective content accessed through the metadata repository

Responsibility: Project Manager; ProQuest; Curtin; Melbourne

Activity	Date
Develop guidelines to identify theses warranting retrospective digitisation	June
Promulgate guidelines	September
Pilot project with ProQuest	June
Commercial agreement with ProQuest	September
Load or link to ProQuest data	December

7.5 Assist in the development of common local site software to handle digital theses

Responsibility: UNSW, Melbourne, Queensland, Curtin, UTS

Activity	Date
Assistance and advice	As required
Test ProQuest online submission system	April

Appendix 1

COMPLIANCE WITH SELECTION CRITERIA

CRITERIA	Yes/No
1. Need for proposal <ul style="list-style-type: none"> • Scope and outcomes • Systems, software used • What will be developed • Timeframe 	Yes Yes Yes in terms of requirements Yes Yes
2. Relevance and significance in terms of Systemic Infrastructure Initiative <ul style="list-style-type: none"> • Impact of using common solutions • Role and importance of open access • Impact of sharing expertise and collaboration • Impact of promoting Australian research output and digital repository initiatives 	Yes Yes Yes (already demonstrated in ADT) Yes
3. Cost-effectiveness <ul style="list-style-type: none"> • Budget • How proposal contributes to enhanced services, skills development etc 	Yes Yes
4. Effectiveness of co-operative arrangements <ul style="list-style-type: none"> • Who is doing what • How bid team involves other institutions 	Yes Yes
5. Sustainability <ul style="list-style-type: none"> • Rationale for a sustainable solution to the need for archiving and preservation • Embedding new practices into existing processes • Potential business models • Various levels of institutional commitment 	Yes (qualified) Yes Yes (in place for ADT) Yes (in place for ADT)

Appendix 2.

ADT central metadata repository exposure and statistics

During the 2002/2003 year, the usage of the Australian Digital Theses Program has increased dramatically. The data below shows that access to ADT has increased 700% and hits by 790% between June/October 2002 and March/June 2003.

PERIOD 1: 30 JUNE 2002 – 31 OCTOBER 2002

During this period, there were **63,484** accesses and **208,697** hits. Each visitor viewed an average of **2.3** pages and downloaded **13.9** kilobytes. Hosts visited **Australian Digital Theses Program** an average of **2.0** times, for a total of **28,198** unique visits of **14,285** unique hosts.

There were **10.2** visits per hour, and a sustained rate of **75.4** hits per hour.

PERIOD 2: 1 MARCH 2003 – 30 JUNE 2003

During this period, there were **359,678** accesses and **1,320,656** hits. Each visitor viewed an average of **10.3** pages and downloaded **72.0** kilobytes. Hosts visited **Australian Digital Theses Program** an average of **2.0** times, for a total of **34,838** unique visits of **17,093** unique hosts.

There were **12.7** visits per hour, and a sustained rate of **482.5** hits per hour.

Appendix 3. Principal contact biographies

1. The University of New South Wales

Andrew Wells, University Librarian. Joined UNSW in 2001. Directed redevelopment of Australian Bibliographic Network to Kinetica at National Library of Australia, 1996-2000. Involvement in standards development with Standards Australia Committee IT/19 and Australian Bibliographic Network.

Tony Cargnelutti, Manager, Online Services Department; project manager of a number of online initiatives, including principal developer and ongoing manager of the ADT Program.

2. Curtin University of Technology

Peter Green, e-Library Development Librarian. Qualifications in education, librarianship and computer science with extensive involvement in the development and provision of innovative online services including; Australian Digital Theses Program at the local and national level; e-Library initiatives such as portals and repositories; and cross-institutional authentication using federated identity.

3. University of Melbourne

Dr. Angela Bridgland, Director, Teaching Learning and Research Support Information Division. Previous position - Director, University of Melbourne Library and prior to that Deputy University Librarian, Uni of Melb. Lib. Seventeen years as a library educator. ALIA - Vic Branch President, Member of Board of Education (2 terms) - currently in second term on the Board of Directors

4. University of Queensland

Heather Todd, Executive Manager, Biological and Health Sciences Library Service, has worked in both special and academic libraries for over 20 years and has a range of experience and expertise in the management of libraries and service delivery. She is presently coordinator of the ADT Program in the University of Queensland Cybrary.

5. University of Technology, Sydney

Alex Byrne, University Librarian. Broad experience in information architecture to store and make appropriately accessible digital information including the design of bibliographic and other databases. Long involvement as an author and editor

of professional writing in information services and in literature with several appointments to Australian and international editorial boards. He also chairs the ADT Policy Group on behalf of CAUL

Appendix 4. ADT extended program – Conceptual Model

