

---

## eScholarship@UQ

### Team Leader

Mr Chris Taylor, Executive Manager, Information Access Services, the University of Queensland Library, University of Queensland QLD 4072.

Tel: (07) 336 56623

Fax: (07) 3365 6888

Email: c.taylor@library.uq.edu.au

### Contact Officer

Ms Belinda Weaver, Co-ordinator, ePrints@UQ, the University of Queensland Library, University of Queensland QLD 4072,

Tel: (07) 3365 8281

Fax: (07) 3365 7930

Email: b.weaver@library.uq.edu.au

## Background and context

Australian universities are creating and, in some cases, hosting increasingly significant collections of their research output in digital form. Yet most of this vast body of information is not easily accessible to the Australian research community. Large collections of it are not indexed by any Web search engine, nor is much of the material openly available to anyone who might wish to see it. Much of the research reported in journal articles and conference proceedings is locked up in proprietary ejournal databases, accessible only to subscribers. Where such material is made available online through the Web pages of individual academics, no clear signposts to this data exist. Researchers also contribute to institutional and national datasets in the social and physical sciences, but again, much of this material is little known, and access to it is sporadic and ill-defined.

In all cases, the mere fact of the material's being electronically available is no guarantee of its being found, used and cited. In the first case, most non-subscribers are locked out of the material, and in the second and third cases, established pathways to such data are lacking. These facts limit and blunt the research impact of much of Australian academia's output at a time when greater online visibility and accessibility is necessary if Australian research is to play a significant role on the world scientific stage. Since wider research dissemination is increasingly possible through electronic means, Australian research institutions need to work together to facilitate this.

While universities must report all research outputs annually to the Department of Education, Science and Training, this valuable reporting is not recorded in a national, searchable dataset that is open to researchers. Accordingly, those who seek information on Australian research must try a variety of mechanisms to identify it.

Piecemeal solutions to building greater accessibility to Australian digital research have encompassed the creation of institutional ePrint repositories in some Australian universities, the Australian Digital Theses Project, and the setting up of ePress initiatives in some Australian universities to create new opportunities for scholarly communication. While these are all worthy projects, their fragmented nature means that researchers must approach each individually and master

different layouts, search facilities and subject schema to discover their contents. The US Mellon-funded OAster project (<http://oaister.umdl.umich.edu/o/oaister/>) does allow cross-searching of many of these initiatives, but searches cannot be searched by subject, which is a limitation that the proposed **eScholarship@UQ** initiative plans to resolve.

Many academics post significant collections of articles, conference papers and other research-related publications on personal web sites, but no clear pathway to any of this data currently exists.

Academic departments and schools also generate large amounts of digital “grey” literature – working papers, discussion papers, seminar presentations and papers, conference posters and technical reports, many of which are hard to access, despite their digital existence. Too often, these valuable records of early research outputs are virtually “lost” to the world; firstly because of a lack of clear reporting and data-gathering pathways even within the home institution; secondly, because of a plethora of differing formats; thirdly, because of a lack of metadata or other structured forms of description; and fourthly because the material is literally scattered across hundreds of different Web servers.

Research centres, institutes and Co-operative Research Centres affiliated to academic institutions also produce a range of digital publications that are not always captured and made searchable by any service other than Web search engines which only do so sporadically and in piecemeal fashion. Hundreds of such publications are produced in formats such as Word, LaTeX and PDF which many search engines do not recognise.

Data on Honours theses and Masters’ theses by coursework is also not captured electronically by many institutions. It is only recently that the content of PhD and research Masters’ theses has been captured and made accessible through the Australian Digital Theses Project, but even this project only captures a fraction of the annual research output of Australian universities. Older thesis information is still of relevance to researchers, particularly in the humanities.

The development of ePrint repositories, such as the University of Queensland’s ePrints@UQ, is a step towards collecting and organising and making searchable research materials from the pre-print stage right through to post-prints and then post-publication updates but these services only capture a fraction of what is published.

A large number of existing research datasets, such as digitised pathology slides, and digital veterinary science images, are also “lost” to researchers because of under-reporting and lack of descriptive material such as metadata. Other such under-reported datasets include digital versions of photographs, sound and video recordings, records of experiments, slide collections, CD-ROMs, DVDs, and software.

## **Project summary**

The **eScholarship@UQ** project will aim to develop an integrated entry point to the vast body of under-reported research, initially within a single institution and then nationally. The gateway will harvest metadata from existing open access repositories, such as those already established for ePrints, and will also establish mechanisms to identify, capture, organise, manage – and in some cases, digitise – other forms of research, such as exists in individual academic Web pages, on departmental or School Web servers and in existing research datasets in both the sciences and humanities, among others. The project’s goals will be to encourage better reporting of academic research outputs, to centralise access to information about research, to improve its visibility and usability, and to add value to existing

information by standardising subject classifications across material harvested from a range of different sources. This would involve mapping existing institutional subject classification schemes to the *Australian Standard Research Classification*. It provides a comprehensive, robust thesaurus specifically designed for Australian research output. Where metadata harvested by **eScholarship@UQ** lacks the appropriate thesaurus descriptors, automatic and semi-automatic subject mapping techniques will be applied during the harvesting and upload process.

## Objectives of the project

The **eScholarship@UQ** project will aim to **develop an integrated entry point to the vast body of under-reported research**, initially within a single institution and then nationally.

At present, it is not clear how research data currently unavailable can best be made accessible. Conversion of data files, content and descriptive elements to XML using automated means appears to provide one solution. The vast plethora of formats and content can more easily then be manipulated for easy access. Automatic addition of metadata tags and the application of a variety of software solutions involving the use of existing ontologies will render the content searchable and accessible.

The project will establish a demonstrator approach to the use of various affordable technologies for conversion of data to XML. The project will also involve the testing of various applications software, emphasising open source solutions, to the resource discovery activity. Database storage aspects will also be investigated.

The gateway will harvest metadata from existing open access repositories, such as those already established for ePrints, and will also establish mechanisms to identify, capture, organise, classify, manage – and in some cases, digitise - other forms of research, such as exists in individual academic Web pages, on departmental, research centre or School Web servers and in existing research datasets in both the sciences and humanities, among others. It will cover electronic journals and theses.

The project's goals will be to

- encourage better reporting of academic research outputs
- centralise access to information about Australian research
- improve its visibility and usability
- add value to existing information by standardising subject classifications across material harvested from a range of different sources.

**The long term aim of eScholarship@UQ** is to establish a single, open access information retrieval resource for all UQ research information. This service will be built through a combination of **harvesting software** (to gather the metadata available) and **storage and retrieval software** (to allow users to search for it). While the service will be centralised for the purpose of search and display, it will work as a decentralised system so that project collaborators can upload data directly to the gateway, should they so wish. The service will be kept up to date by regular re-harvesting of existing repositories or by direct record input from project partners. While the gateway will provide a single gateway for searching purposes, material may be held in a range of different databases, accessible through an integrated search function.

---

The primary criteria for inclusion in the eScholarship@UQ service are that the information resources the metadata describe

1. originate from Australian research bodies, such as universities, research centres and institutes
2. are publicly accessible and have no access restrictions on full content
3. have a corresponding Web-based digital representation, or be capable of being digitised

The proposed service will encompass as broad a collection of resource types as possible within Australian research, and will actively seek new materials for inclusion within the service.

## Platforms

In order for material in **eScholarship@UQ** to be as accessible as possible, thus guaranteeing the widest possible research impact worldwide, all software used in the project will be open source. **eScholarship@UQ** will be a Web-accessible set of databases, built upon the Linux operating system, using a MySQL database with Perl and PHP as the programming languages. It will comply with a number of Australian and international standards, such as the Open Archives Initiative, for open access collections; enhanced Dublin Core, for metadata; the OpenURL standard, for long-term linking and access; and the Australian Standard Research Classification, as a subject thesaurus. The use of XML as a format for data exchange will also be investigated.

## Collections

- The **eScholarship@UQ** gateway will target collections of electronic research papers, such as journal articles, working papers, conference papers, posters or proceedings, technical reports, discussion papers, books and book chapters, slide collections, image libraries, data on experiments, online journals, e-books, honours and coursework masters' theses, preprints and other types of research output such as software programs
- PhD and Research Masters theses, such as those indexed by the Australian Digital Theses project
- Academic e-journals
- Other collections identified during the course of the project

## Expertise

The University of Queensland Library has considerable expertise in creating and managing digital collections. The Library has participated in a number of nationally significant projects such as the Australian Digital Theses project, the AVEL Sustainability Network, the WebLaw initiative (where UQ was the lead institution), ePrints@UQ (an open access repository of UQ research), the DigiLib project (architectural images of Queensland). The Library is also a contributing partner to PictureAustralia. The Library also has considerable expertise in Web-database integration programming, system administration and hosting, metadata, including an existing metadata repository code library (developed by UQ Library Technology Service staff for the WebLaw project) and in programming languages such as Perl, PHP, Java and Javascript.

The development of various database structures for data storage and the use of appropriate indexing with various existing projects will be extended.

## eScholarship@UQ

### Budget

<b>Hardware</b>	Web server and storage	\$15,000.00
<b>Web hosting, administration and system maintenance</b>	Including domain registration, mail, etc	\$10,000.00 pa
<b>Software development</b>	Harvesting software	\$15,000.00
	Subject mapping	\$15,000.00
	Repository design /adaptation	\$10,000.00
	Graphic and Web design	\$5,000.00
	OpenURL linking	\$10,000.00
<b>Staffing costs</b>	Scholarly Communication Director/Research Officer	\$100,000.00
	Data Analysis / Research Officer	\$80,000.00
	Project co-ordinator	\$50,000.00 pa
	Project liaison	\$10,000.00 pa
	Collection identification and follow-up	\$5,000.00 pa
<b>Marketing, publicity, mailings</b>		\$5,000.00
<b>Travel costs / scholarly communication liaison</b>		\$25,000.00
<b>Total:</b>		<b>\$355,000.00</b>

### What UQ can offer in kind

- An existing metadata repository code library (developed by UQ Library Technology Service staff for the WebLaw project) that could be adapted for use in **eScholarship@UQ**
- Hosting, administration and maintenance of the **eScholarship@UQ** service