

Summary of UQ Testbed Program for APSR

The University of Queensland proposes to investigate the feasibility of “local harvesting” of born digital and not yet digitised UQ research material as part of its testbed program for APSR. The aim of such local harvesting is to allow UQ to make available large amounts of data that would be of use to researchers, were it possible for the data to be captured, stored in a central repository (such as UQ’s existing **ePrints@UQ** repository), fully described with appropriate metadata, and made searchable.

Many UQ 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. In addition, academic departments and schools at UQ 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 UQ 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 currently captured electronically. 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. 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.

The UQ testbed project will seek to identify all such material and look at the feasibility of capturing it and making it digitally available.

23 March, 2004