

Bibliographic Control of Australian Higher Degree Theses: The Future Role of the Australian Digital Theses program

PAUL GENONI, and ROBERTA COWAN

ABSTRACT Australia makes a substantial investment in the production of higher degrees by thesis. Completed theses form an important body of original research data, and underpin the career of many of the nation's leading researchers. There is evidence, however, that there is inadequate bibliographic control of this material, and that as a consequence important research data may be difficult to locate. This article traces attempts to provide access to bibliographic records of Australian theses, highlights the problems with regard to both accessibility and quality of those records, and suggests that these problems might be addressed by an expanded role for the Australian Digital Theses program.

Higher degree (PhD and Masters) theses completed in Australian universities are an important component of the nation's research output. They distil the results of research conducted under supervision across the full range of disciplines, and although not 'published', they have undergone a rigorous form of peer-review courtesy of the requirement that they are examined by recognised subject experts external to the degree granting university. Many of the country's most esteemed academics and researchers have based their careers on the research undertaken for their higher degrees.

Australians as a whole also stand to benefit from the information and knowledge embedded in the corpus of completed theses. Not only do theses comprise a history of an important component of the nation's research activity – in many cases carried out under circumstances and conditions that can never be repeated - but they form an extensive body of research data crucial to attempts to forge a knowledge based economy.

It is therefore an issue of considerable importance that these documents are adequately recorded in easily accessible and searchable bibliographic sources. This task has fallen largely to the nation's university libraries, and although considerable effort has been expended, there is evidence that it has not been wholly successful.

Investment in research training

The Australian community, through the offices of the Commonwealth government, makes a substantial investment in the production of research leading to higher degree completions. Allocation by the Department of Education, Science and Training (DEST) for the support of the Research Training Scheme in 2003 is \$527,399,000, made to 41 Australian higher degree awarding institutions. DEST will provide a further \$87,126,000 by way of scholarships under the Australian Postgraduate Award Scheme, and \$17,302,000 through the International Postgraduate Research

Dr Paul Genoni, Lecturer, Department of Media and Information, Curtin University of Technology, Perth, GPO Box U1987, Perth 6845. Email: p.genoni@curtin.edu.au

Dr Roberta Cowan, Adjunct Lecturer, Murdoch Library and School of Biological Sciences and Biotechnology, Murdoch University, South St, Murdoch, Perth, Western Australia, 6150. Email: cowan@possum.murdoch.edu.au

Scholarships for international students to undertake higher degree research at Australian universities.¹ In addition these students call upon the time and skills of their supervisors, who represent the most qualified and experienced researchers working in Australian universities.

The number of students undertaking and completing higher degree theses has been growing rapidly. In 1991 there were 19,254 enrolments in higher degree by research programs in Australian universities, and by 2000 this had risen to 37,374. Across the same period higher degree by research completions rose from 2491 to 5325. A total of 26,620 PhD and 15,454 Masters degree theses were completed during the decade from 1991 to 2000.²

Part of the substantial investment in higher degree research is recouped through the process of 'credentialling', whereby graduated PhD students in particular form the basis of the next generation of teachers and researchers in public and private institutions. Their degree is not only evidence of their skills but also a critical part of the process they undertake in order to acquire those skills.

The theses themselves, however, also have a considerable ongoing educational, social and economic value to the country. They form a rich source of original research data covering the entire spectrum of disciplines and subjects. Some of this data is shared with research communities when it is published in journal articles or books, but it is also the case that much of it remains unpublished or published only in an abbreviated and fragmented form. This is increasingly the case in those subject areas - particularly from the humanities - that have traditionally depended upon academic monographs in order to distribute research results. The ongoing decline of the academic monograph, coupled with the ever-growing number of theses being completed, means that full-length publication has become the exception. For a growing number of students publication or distribution of the key research outcomes in any form other than the thesis itself is increasingly unlikely.

Universities and their libraries have responded to the need to record and make accessible completed higher degree research by collecting a printed copy of theses (PhD, Masters and in some cases Honours) submitted at each institution. In this way, there is at the very least a copy of the thesis in the library of the institution awarding the degree, and a record of that holding in the local catalogue.³

The problem

Despite the work done by libraries in collecting and recording theses, access to this material currently suffers from the absence of a single, easily accessible bibliographic source providing comprehensive coverage of completed theses.

The foundation publication in this regard was the *Union List of Higher Degree Theses in Australian University Libraries*. The *Union List* was produced and published by the University of Tasmania Library and first issued in 1959. The initial publication was supplemented by a number of revisions and annual compilations, with the final volume being issued in 1991, but including coverage to 1989 only. For this substantial period, however, the *Union List* was the most important bibliographic tool covering Australian theses.

The *Union List* had in part been made necessary by the lack of coverage given to Australian theses by international bibliographies – a situation that continues to this day. The major international source for bibliographic data relating to theses, *Dissertations Abstracts International*, includes entries from a small number of

Australian universities only, although it should be noted that it includes numerous references to theses of Australian relevance produced at non-Australian universities.

The *Union List* was an important location tool but it provided limited subject information. For although it was organised in a classified sequence with a keyword subject index it did not include abstracts – a vital element in assessing the relevance of documents which necessarily deal with complex subjects. The issue of whether to include abstracts in the *Union List* was a decision which, according to bibliographer Dietrich Borchardt, was ‘discussed at great length among librarians and university authorities and in spite of several requests for the inclusion of abstracts or explanatory notes the consensus view was not to burden the editors with the task of preparing such abstracts’.⁴

The cessation of the *Union List* in 1991 was announced by the University of Tasmania Library in the following manner:

... history and technology have overtaken us in this enterprise and in the future, the responsibility for collecting and disseminating information on Australian higher degree theses will rest with the National Library of Australia, with data on theses recorded on ABN (the Australian Bibliographic Network). The National Library will publish supplements to replace this series.⁵

Very shortly after the demise of the *Union List* the issue of bibliographic and physical access to theses received some attention in the lead up to the Towards Federation 2001 conference, held in March 1992. A Working Paper (no 14) was prepared, highlighting the ‘problems in the collection, preservation, bibliographic control and access to a large proportion of the nation’s theses’. The problems identified included the poor state of some records provided to the *Union List* which meant that they could not be loaded to ABN in their existing form; the inadequate subject access that existed without abstracts or some other form of enhanced subject description; and the need for a very high level of physical description which reached ‘at least full AACR2 level 2 cataloguing’.⁶ As with the University of Tasmania Library, the Working Paper envisaged a situation whereby the National Library would undertake the preparation in printed form of an annual list of completed theses.

Despite the issues raised by the Working Paper there is no evidence in the records of Towards Federation 2001 that the matter of theses was discussed at the conference. It is likely that this was because it was widely believed that the National Library would be responsible for continuing to publish a revamped version of the *Union List*, or that bibliographic access to records via ABN would be sufficient. Unfortunately, the National Library never issued any supplements to the *Union List* or other form of printed guide to theses, and numerous records of items that were included in the *Union List* were never converted for inclusion on ABN. Just as troubling, not all Australian universities contributed cataloguing data to ABN and later Kinetica.⁷ Therefore the effort to provide a single comprehensive record of Australian theses died with the *Union List*.

For other reasons also, ABN and Kinetica have proven to be frustrating sources of information about theses. Limitations in the search capacity of the two systems, coupled with inconsistencies in cataloguing practices between institutions, means that quite sophisticated search skills are required in order to isolate theses from other types of documents. As the DEST Information Infrastructure Advisory Committee recently reported, ‘Although most universities include records of theses in the Australian National Bibliography, they are not easily searched as a set and there is no means of locating all theses in a particular discipline’.⁸

As the Towards Federation 2001 Working Paper indicated, the problem of separating theses within ABN/Kinetica – or indeed other forms of union catalogue – is, in part, the result of the inadequate and inconsistent forms of cataloguing which have been applied. Standards for subject indexing, including the *Library of Congress Subject Headings*, do not provide a standard heading or subheading by which theses can be identified, and institutions have therefore frequently devised their own solution to the problem of identifying these items. This may be by the use of local subject headings or by an addition to a record using a ‘notes’ field. A survey of current practice in university libraries shows that catalogue users are directed to search using various keywords including ‘thesis’ (for example, the University of Western Australia), ‘theses’ (Deakin University) and ‘dissertations’ (Curtin University of Technology).

There is also evidence of uncertainty within university cataloguing departments as to how theses should be described. For obvious reasons, cataloguing these documents requires original records to be created, and this has led to error or inconsistency. Again, even a brief survey of university library records reveals frequent variations in practice within institutions.

In an effort to provide improved bibliographic access to theses a number of specialised bibliographies have been prepared over the years. Some were issued in serial form but many were simply ‘one-off’ publications that provided a snap shot of the titles available at a point in time. They were also limited in the breadth of their subject coverage and irregular in the manner in which they selected ‘Australian’ theses. For example, some attempted to provide coverage of all Australian theses on a subject (e.g. *ASBS Thesis List*, Canberra: Australian Systematic Botany Society, 1978, plus supplements); others provided lists of Australian and international theses on topics relevant to Australia (e.g. Malcolm Tull ed., *A Bibliography of University Theses on Australian Maritime History*, St Johns: International Maritime History Association, 1996); and yet others included Australian sourced theses on essentially non-Australian subjects (e.g. Frank Shulman ed., *Doctoral Dissertations on South Asia 1966-1970: An Annotated Bibliography Covering North America, Europe and Australia*, Ann Arbor: University of Michigan, 1971). Although these publications have undoubtedly been useful to researchers, they have done little to compensate for the lack of a comprehensive bibliography of Australian theses.

Australian Digital Theses

More recently considerable international attention has been given to the desirability of having theses available electronically in full text. This work has been led from the Virginia Polytechnic Institute which has underpinned the development in the USA of the Networked Digital Library of Theses and Dissertations (NDLTD). The NDLTD initiative has been widely influential, and Australia has responded by implementing the Australian Digital Theses (ADT) program. According to the ADT the program ‘is designed to improve access to, and enhance transfer of, the research information contained in theses by providing a full text version available from the desktop via the web’.⁹

The initial development of the ADT was funded by a 1997 grant from the Australian Research Council to seven Australian universities operating under the auspices of CAUL. Membership was opened to other CAUL libraries in 2000, and as at August 2002 twenty-six universities were members of the ADT. To date, however, only a comparatively small number of theses (848 as at 7 March 2003) has been made available from member institutions.

Since its inception the ADT has become recognised for its innovative structure. The program is based on a distributed database, with member institutions undertaking to load deposited theses onto local servers in PDF format, and load the metadata which provides bibliographic access to the theses on a server at the University of New South Wales. The metadata is generated by research students when they elect to deposit their thesis by submitting a form made available from ADT members.

In the development of standard metadata fields the ADT has made some advances in the provision of bibliographic access to theses. As indicated above, the recording of theses has previously been characterised by error and inadequate and inconsistent levels of both physical and subject description. The ADT has served to standardise the elements of the physical description, with a series of fields derived from the Virginia Polytechnic. The reliance on the author to provide this detail, however, will inevitably lead to inconsistencies, particularly in the form of names applied to supervisors or academic departments. If students report the name of a department incorrectly, this seems in many cases to be transferred in this incorrect form to the cataloguing record. While this problem can be overcome by careful authority control, there is a second and far more intransigent problem in creating an accurate record of academic departments. That is, the frequent change of names of departments created by either the realignment of academic units within universities, or by their need to reflect new terminology and emerging areas of study.

Subject description also remains problematic, with no attempt made to impose a standardised form of indexing. Title and author-generated keywords are the principal points of subject access, although ADT records are enhanced by the addition of searchable author-generated abstracts. The ADT also has an advantage over Kinetica in having an interface that is user friendly and easily managed by searchers with even a moderate level of familiarity with electronic information retrieval.

The ADT will undoubtedly continue to grow and become an increasingly important resource, but it is highly unlikely that it will reach the point where it provides comprehensive coverage of recently completed theses. Universities still show a strong preference for the use of print as the preferred mode for storing archival copies, and the contribution of digital copies to the ADT is likely to remain optional for students of some (perhaps most) member institutions. The universities continue to have concerns about the same issues of long-term preservation that hinder many other digitisation projects, and some students are worried about the integrity of digitised text and the diminished possibility of secondary publication of research results that are made available so freely.

The ADT database will therefore in all likelihood remain fragmentary, even for recently completed theses, and its primary benefit will be to make the included theses far more accessible. Indeed, one recent USA calculation (based on data from the Virginia Polytechnic Institute) concluded that digital theses 'are 100 times more likely to be circulated than print theses and dissertations'.¹⁰

Despite the achievements of the ADT, the problem remains the lack of a single source of comprehensive and readily obtainable bibliographic coverage of Australian theses. It is an issue that is significant enough to cause concern for a number of university libraries. For example the University of Adelaide Library webpage warns users that 'Australian theses are not well indexed after 1985 (sic)';¹¹ the Monash University Library states that 'There is no comprehensive list of all theses submitted in Australia',¹² and Deakin University Library declares that 'It is not always easy to locate details of theses nor to obtain a copy of a thesis once identified'.¹³

Current planning

It seems apparent that any changes to the current situation with regard to the bibliographic control of theses are likely to be only achieved within the ADT framework. The current funding environment heavily favours digitisation projects and as the ADT proudly boasts in its current business plan it is the 'only game in town' when it comes to the digitisation of Australian theses.¹⁴

In the ADT *Business Plan 2002-2006*, CAUL has identified priorities for the ongoing development of the program. These include ensuring as comprehensive as possible coverage of newly completed theses. This will be undertaken by 'achieving total participation by CAUL members', and by 'advocat(ing) electronic submission of theses as the preferred mode of submission'.¹⁵ It is quite likely that the former will be achieved, but for reasons stated earlier, the latter is far less likely.

The ADT plan also aims to 'develop content through retrospective conversion of frequently requested or notable theses'.¹⁶ Clearly, however, this is an expensive task. It is one of the advantages of the ADT's distributed nature and author-generated submission, that running costs for the program are minimised. CAUL allocated a modest \$75,000 for the operation of the program in 2002, with additional costs being met by the member institutions. Any significant attempt to digitise older theses would entail a substantially larger commitment from CAUL and/or the universities.

The cause of retrospective conversion has been taken up, however, in a November 2002 report by the DEST Information Infrastructure Advisory Committee (IIAC). The Committee has identified 'Digital theses' for priority attention as one of 'six areas of strategic importance relating to the creation and dissemination of Australian research information'.¹⁷ It envisages a 24 month project with a series of aims intended to expand the content of the ADT, including 'identifying key theses (individually or in discipline groups) for priority digitisation', while acknowledging that 'experience, however, suggests that retrospective digitisation of all theses to be impractical'.¹⁸

A survey of the ADT database conducted in March 2003 reveals that its content is – not surprisingly – dominated by recently completed theses.

Pre 1970	1
1970-1979	5
1980-1989	17
1990-1999	231
2000-2003	591 ¹⁹

A conservative estimate indicates that there are in excess of 90,000 higher degree theses of Australian origin not on the ADT database. Notwithstanding that many of the more recently completed of these theses already exist in a digital form, the task of comprehensive conversion would nonetheless be well beyond any resources likely to be made available. The IIAC has indicated a 'notional allocation' of \$800,000 - \$950,000 for their envisaged project, including 'at least \$500,000 to assist institutions to make content available'.²⁰ There is no indication as to how many theses might be retrospectively converted with this amount of money.

The other aims for the future development of the ADT that would be achieved with an injection of the suggested sum of money are outlined only very generally in the IIAC report. It is, however, a feature of the IIAC report - and something not suggested by the ADT *Business Plan 2002-2006* - that it may be desirable to extend the Program in such a way that theses could be represented in forms other than full text. The IIAC raises the possibility of the first phase of the recommended project 'advising on the

relative merits of maintaining a database of only full-text digitised theses, abstracts only with digitisation on demand or a mix of approaches'.²¹

A future role for the ADT?

This suggestion by the IIAC seems to point the way towards at least a partial solution to the problem of providing better bibliographic access to Australian theses. Indeed, it would seem to be a natural extension of the ADT that it becomes a repository for all extant bibliographic data relating to completed theses. Currently, a researcher wishing to undertake a search for theses would need to search at least the *Union List*, Kinetica and the ADT database in order to ensure reasonably comprehensive coverage. Even this would not guarantee that all relevant theses had been located, and it would require a highly skilled, informed searcher – a librarian with some familiarity with all three sources - to obtain anything approaching a complete result. The task is further complicated if a researcher wishes to locate not only theses of Australian origin but also of Australian relevance, in which case the search should be extended to include at least *Dissertation Abstracts International*.

It is suggested that this problem should be addressed by aggregating the bibliographic data relating to theses of Australian origin and Australian relevance on the ADT database. Given the impracticality of retrospectively converting the text of all theses, this would require the aims of the Program to be modified in order to accept bibliographic records without the full text attached. That is, some records would reside on the database with metadata only.

To make an enhanced ADT database as complete as possible in its bibliographic coverage of relevant theses would require the following:

- isolation of all relevant records on Kinetica, and the transfer of these records to the ADT;
- identification of items listed in the *Union List* which have not been converted and migrated to Kinetica, and the creation of records for inclusion in the ADT database;
- identification of theses that are not represented in either the *Union List* or Kinetica, and the creation of records for inclusion in the ADT database;
- provision by universities which would not otherwise supply a copy of the text of a thesis to the database (either because of institutional non-membership of the ADT or a student's refusal to grant permission) of a full bibliographic record including abstract. An alternative would be for the ADT to continually harvest metadata from university catalogues or Kinetica – although this would ideally require agreement on standards, including the addition of abstracts, for future cataloguing;
- identification and addition of records for theses of Australian relevance but of non-Australian origin. The records of *Dissertation Abstracts International* would provide a valuable starting point for this task.

There would be inconsistencies in the amount and quality of metadata attached to these different categories of records, although this problem would be mitigated in that many records from the *Union List* have already been added to Kinetica with upgraded subject headings attached. Nevertheless, a database modified in this way would inevitably include some records with access to title only; some with title plus library added subject headings; and still others with title, author-generated keywords and abstracts.

Given the difficulty of providing adequate subject access to theses using standard lists of subject headings, the most rewarding way to further enhance the usefulness of the

database would be to focus on the addition of abstracts. Ideally, these should be in a format that makes abstracts available for searching, as with those currently on the database. If this proves to be too expensive, then abstracts should be added as scanned non-searchable images. It is a fortunate characteristic of theses that they frequently have lengthy, descriptive titles including many keywords that would otherwise be found in author-generated subject headings or abstracts. Therefore, even those theses entered with title as the only point of subject access are still likely to be retrieved. In these circumstances an abstract, albeit non-searchable, will nonetheless add information valuable to the searcher.

The problem of subject access that is not wholly controlled or standardised would, however, be substantially outweighed by the benefits of having bibliographic records of all theses available in one database with an interface suitable for end user searching. In this proposed form the ADT database would allow searching on only those records which had the full text attached, or on all records.

It is therefore suggested that funds provided for further development of the ADT database be used in this way, rather than for the conversion of the full text of theses. Full text conversion should be undertaken only when strong demand might be established by a large number of hits on a particular record. Subsequent developments would include the gradual upgrading of incomplete or inaccurate records.

It is also suggested that the National Library should address the issue of standards for the cataloguing of theses. It is apparent that although most libraries are attempting to follow the current international standards, that these are being interpreted differently and are subject to local policy and practice. An Australian standard that is compatible with the developing requirements of the ADT would be a useful innovation.

Conclusion

This expanded role for the ADT would render the *Union List* obsolete, and replace Kinetica as the primary source of information about Australian theses. Australian and international researchers would, for the first time, have comprehensive access to the bibliographic record of Australian higher degree research, and of higher degree research related to Australia from non-Australian sources. This improved coverage would be coupled with enhanced levels of subject retrieval and subject information about individual theses.

¹ Department of Education, Science and Training *Higher Education: Report for 2003 to 2005 Triennium* Canberra 2002
http://www.dest.gov.au/highered/he_report/2003_2005/pdf/triennium2003_2005.pdf [accessed 24 February 2003]

² Department of Education, Science and Training *Students 2001: Selected Higher Education Statistics* Canberra 2002 <http://www.dest.gov.au/highered/statistics/students/01/tables.htm> [accessed 24 February 2003]

³ It is probable that a number of theses are not deposited as required. Individual universities are responsible for ensuring that deposit occurs, but in some cases at least compliance is not supported sufficiently by regulation or procedure.

⁴ D H Borchart *Australian Bibliography* Melbourne Pergamon 1963 p34

⁵ University of Tasmania Library *Union List of Higher Degree Theses in Australian Libraries: Supplement 1989* Hobart University of Tasmania Library 1991 p vii

⁶ Towards Federation 2001: Linking Australians and their Heritage: A National Conference on Access to Australia's Recorded Documentary Heritage *Final Report: To Which is Attached Agenda Papers, Working Papers, Background Papers* Canberra National Library of Australia p205

⁷ A guide to the coverage of theses on Kinetica for the various universities current to 1998 can be found at National Library of Australia *Australian Higher Degree Theses on the National Bibliographic*

Database <http://www.nla.gov.au/abn/covrthes.html> [accessed 24 February 2003] CAUL conducted a survey in 2002 to update this information, but the results are not available at the time of writing.

⁸ Department of Education Science and Training Information Infrastructure Advisory Committee *Research Information Infrastructure Framework for Australian Higher Education* Canberra 2002 p38

⁹ Australian Digital Theses *Aims & Overview* <http://www.library.unsw.edu.au/thesis/adt-ADT/info/aims.html> [accessed 5 March 2003]

¹⁰ J Moxley 'Universities should require electronic theses and dissertations' *Educause Quarterly* no 3 2001 p61

¹¹ University of Adelaide Library *Dissertations and Theses by Place of Origin* <http://www.library.adelaide.edu.au/gen/theses/otheraus.html> [accessed 25 February 2003]

¹² Monash University Library *How to Find a Thesis* <http://www.lib.monash.edu.au/vl/thes/thes06.htm> [accessed 25 February 2003]

¹³ Deakin University Library *Theses and Dissertations* <http://www.deakin.edu.au/library/thesdiss.html> [accessed 25 February 2003]

¹⁴ Council of Australian University Librarians *Australian Digital Theses Business Plan 2002-2000 Revised Version September 2002* <http://www.anu.edu.au/caul/adt/adt-business-plan.doc> [accessed 4 March 2003]

¹⁵ *ibid*

¹⁶ *ibid*

¹⁷ Department of Education Science and Training Information Infrastructure Advisory Committee *Research Information Infrastructure Framework for Australian Higher Education* p18

¹⁸ *ibid* p39

¹⁹ The total number of dated theses (845) is slightly lower than the total number (848) reported on the database as at 7 March 2003. Several theses were found with no entry in the date field.

²⁰ Department of Education Science and Training Information Infrastructure Advisory Committee *Research Information Infrastructure Framework for Australian Higher Education* p39

²¹ *ibid.*