A New Continent into Literature: The Australian Literature Database at the University of Sydney Library

Dr. Creagh Cole
University of Sydney Library
c.cole@library.usyd.edu.au

Introduction
In 1973, the Australian novelist Patrick White was awarded the Nobel Prize for literature for his novel “The Eye of the Storm.” At the ceremony in Stockholm the novelist was honoured for his “epic and psychological narrative art, which has introduced a new continent into literature.” This was undoubtedly true and White’s achievements have been followed by a new generation of novelists, poets and dramatists: Peter Carey, David Malouf, Les Murray, to name only a few. Patrick White’s award has always seemed to Australians to have tremendous significance, perhaps because it put to rest internationally the old colonial worries about whether Australians were capable of producing truly great art. However, there has always been a problem for contemporary Australian literary studies in its lack of any sense of a tradition predating our modern writers. It is as if we Australians did not begin to write about our experiences until the middle of this century. There were, of course, important writers before and contemporary with Patrick White, and he himself named a number as worthy of the acclaim surrounding his own work. But leaving aside the debates and concerns about what constitutes “great art” and “great literature”, there is a more “archaeological” approach to a nation’s literary output, and one that is peculiarly amenable to electronic forms of publication and dissemination. This approach is evident in the creation of a database at the University of Sydney Library - The Australian Literature Database: A Collection of 18th 19th and 20th Century Australian Texts.

The written text in Australian history is a record of successive waves of European migration and adaptation to an utterly new and different geographic space. It is a record that begins in the journals, letters and diaries of explorers, early settlers and travellers, and develops with the emergence of truly literary forms around the 1840’s: novels, plays, verse, short stories, critical works and more. This is a unique record of the engagement of new settlers with a bewildering new environment and of the emergence of a distinctively Australian voice and sense of identity. It is a record of personal hopes, of loss, of tragedy, of dry humour, of deep disappointment and of great triumph. There is much to be proud of in this record and, regrettably, much of which to be ashamed. It is a record of where we have come from, of where we are, of who we are, and it is a record with which Australians as a whole should be familiar in these times of national re-assessment. It is, or should be, a record of some significance to non-Australians as well.

And yet it is a record that remains relatively obscure and unknown, even within Australia, much less internationally. The key to unlock the treasures of this unique body of work, in my view, lies in the digitisation and electronic publication of a significant number of the key texts. Widespread access to these texts through electronic delivery, texts otherwise available only in the nation’s major research collections, has the potential to radically alter research in Australian
For the first time, researchers will be able to track effectively the emergence of particular terms, concepts, and influences in Australian culture over time, or to track the differential impact of the new environment on men and women writers. The work of creating such a collection of texts has been started at the University of Sydney Library.

SETIS at the University of Sydney Library

The Scholarly Electronic Text and Image Service (SETIS) was formed at the University of Sydney Library in late 1995, with the aim of supporting textual studies at the university. Its creation coincided with a short visit by David Seaman in 1995 and we were persuaded that a possible direction for the service was in creating our own texts, using the infrastructure to be built to network our existing collection of sgml-encoded electronic texts. Adapting scripts supplied by the University of Virginia Electronic Text Center, by the end of 1996 we had achieved some hard won experience in networking for the web such diverse full text collections as the English Poetry and the Patrologia Latina databases. In July of 1996 I attended the Center for Electronic Texts in the Humanities (CETH) Summer School conducted at Princeton University and had the good fortune to learn about the Text Encoding Initiative (TEI) in classes conducted by Michael Sperberg-McQueen. By the end of 1996, we had the technical infrastructure and software in place and were developing some familiarity with TEI for the encoding of humanities source texts. We were also experimenting with scanning equipment and software to prepare digital files of significant local library and archival materials. In short, the library had a platform from which it could launch its own electronic publishing projects.

One of the first projects was a collection of lecture notes by John Anderson, the Scottish libertarian and critical realist philosopher who held the chair of Philosophy at Sydney between 1927 and 1955. Professor Anderson was an extraordinarily influential character, not only at the university but also within Australian cultural and political life generally. Students and researchers from all over the country regularly consult these lecture notes, most hand-written. Our aim was to key these handwritten notes in order to provide a fully searchable text file linked to scanned gray-scale images of each page for comparison. A second project presented itself when the Special Collections staff members at the library were requested for copies of the New Australia newspaper of the 1890’s. This was a publication associated with the utopian socialist expedition of the period to establish a colony in Paraguay led by the Queensland socialist labour leader William Lane. The newspaper itself was in an advanced state of deterioration and presented an ideal test case for the creation of a digital version. Two further advantages flowed from this project. The first was that we were able to create a virtual collection based on the print holdings of several Australian research collections, not just our own but also those at the State Library of New South Wales (housing the major Australian research collection, the Mitchell Library) and at the City of Sydney Library. The second advantage was that we gained experience in the EAD (Encoded Archival Description) and EBIND (Electronic Binding) dtd’s to prepare finding aids to the collection of digital images we were creating, and to manage the expanding collection of images itself.

Other projects have been initiated by academic staff at the university and include: a collection of about twenty Italian medieval one-act plays; the first edition of “How Labour Governs”, a classic work in Australian labour history from 1923 by the Oxford historian and archaeologist Vere Gordon Childe; Mary Shelley’s manuscript autobiography of her father, William Godwin; important historical and current botanical works on Australian eucalypts and forest flora.

Australian Literature: The Collection

Our largest project, however, emerged in 1997 as a result of our involvement with a major digital conversion project led from the University of Sydney Library and the National Library.
of Australia. The Australian Cooperative Digitisation Project 1840-45 (ACDP) aimed to create
digital images of newspapers and journals for the period 1840 to 1845. It also included, in
addition to these newspapers, four novels from the period, which were selected by academic
advisers as seminal works in the emergence of a distinctly Australian literary tradition. These
works were:

- Mary Therese Vidal. Tales for the Bush (1845)
- Thomas McCombie. The Adventures of a Colonist (1845)
- Charles Rowcroft. Tales of the Colonies (1843)
- William Christie. Love Story by a Bushman (1841)

The aim of the ACDP leaders was to prepare digital images of each of these novels and process
these images to prepare ascii text files. SETIS offered to encode the resulting files according
to the TEI guidelines and to publish them at the web site. The natural consequence of the
completion of this work was to seek to extend the collection from that five-year period to
encompass all of nineteenth century literature. The result of such an undertaking would be the
ability to search keywords and phrases across various date ranges, in this way tracking the
emergence and evolution of significant concepts in Australian history and culture. The ultimate
aim of the project would be to achieve comprehensive coverage based upon the major
bibliographic works in Australian studies: Ferguson’s Bibliography of Australia, and Miller’s
Bibliography of Australian Literature. In the meantime, an advisory group was to be estab-
lished to advise on a manageable number of key texts to be added to the collection, as well as
other decisions to be taken about the database, which may affect its research utility. Members
of this advisory group include academic staff in Australian literature from the Universities of
Sydney, Queensland and Western Australia, as well as from the Australian Defence Forces
Academy in Canberra, which has long been a centre for the study of Australian literature.

The immediate problem, however, was to increase the size and coverage of the collection by
searching out any electronic files that may already exist. Some of these were found on the web
at various sites in Australia and overseas. Others were donated by institutions such as the
Mitchell Library (at the State Library of New South Wales); the Macquarie Dictionary Ozcorp
database (scanned texts used in the creation of the standard Australian dictionary); and the
Scholarly Editions Centre at the Australian Defence Forces Academy (ADFA). The result is a
collection of around 100 texts, already a sizeable database of full-text literature of this kind.
The collection is indeed, the largest electronic collection of Australian source texts and already
includes the major works to be found in any good anthology of the literature.

The collection has grown almost by accident. There has been no overall plan or systematic
attempt at collection building. Because the early collection of material relied upon several
sources, the quality of the digitisation of the texts has been variable. Many of the texts have
been double keyed and have a high degree of accuracy. Others have been scanned, and proces-
sed using OCR software to produce the text file for encoding. Each file produced in this way
has been through at least one correction process, most more than this. The number of times
each text has been proofed varies, although each time is recorded in the header information. In
many cases, compressed tiff images of each page are provided with the text file, as well as gif
and jpg images of book illustrations. It would be desirable to have quality high resolution
images available for all of the texts, but particularly for the earlier periods. The encoding of
each text has been fairly light, except with regard to the TEIHeader, which allows us to
incorporate a large amount of bibliographic, and other information about the electronic file
and the printed source text. Within the body of the text, besides basic structural divisions and
headings, features appropriate for the different kinds of literature are marked, such as speakers
and stage instructions in dramatic works, lines of verse, highlighted text, regularised versions of particular terms and spellings, footnotes, page numbers from the source text, passages of spoken dialogue in prose works, quoted passages from works external to the text, and so forth. We are seeking national government funding to continue the project, as this will allow us to follow more closely the guidelines established by the academic advisers to the project, and, eventually, to realise the distant aim of providing a full-text database corresponding to the major bibliographic works listed above. It would also allow us to give more attention to the production and provision of digital images corresponding to the text collection.

Although the collection has grown largely without systematic guidance in the selection of works to be included, it is a unique collection, and we should not underestimate its current charm and strengths as a selection of significant literary, historical and political works. The collection displays great diversity. Australian literature is represented in the collection in all its forms: novels, plays, autobiographies, poetry, explorers’ journals, and so on. The texts include early explorers’ journals such as Sir Joseph Banks’ journal of his voyage with Cook on the ship “Endeavour”, and the Prado manuscript of the early Portuguese encounters with the continent; early historical accounts of the first settlement at Sydney Cove by Watkin Tench; novels, plays and verse of the convict and gold-rush eras; works by the popular nationalist writers of the 1890’s like Henry Lawson, A.B. Paterson, and Steele Rudd. The search for existing files has retrieved a large number of these more popular writers, but the collection does not neglect the lesser-known authors of the period. This will be perhaps the enduring value of the project: the republication of works otherwise unavailable to the general reader. The collection includes a large number of works by women authors and poets. Readers will find in the collection the great convict novels and stories of John Lang, Marcus Clarke and Oline Keese, the goldrush novels and diaries of Mary Gaunt, H.H. Richardson, and the “foreign anarchist” Carboni. They will be reminded that many more nationalities settled the country following the convict era than the British and Irish first settlers. Bush stories and humour abound in this collection, reflecting a traditional Australian identification with the country rather than the city, but also represented are a range of urban-based literary works and political texts which is only appropriate in one of the world’s most urbanised cultures. Even amongst the bush stories there is great diversity, ranging from the affectionate, populist humour of Steele Rudd, to the more reserved, darker humour of Edward Dyson and Henry Lawson, and on to the more disturbing, often frightening portrayal of the bush by Barbara Bayton. At all levels there is great diversity in this collection.

There are two basic advantages in the electronic publication of these texts: the research functionality provided for the texts, and the simple matter of greater access to the texts by a very wide audience within and without Australia. One of the important aspects of the collection as it is now conceived is the ability to search the database according to gender, literary type, time period and so forth. The kinds of research allowed for by the database will change over time as researchers engage with the collection and probe its ability to answer their questions. There is already, however, a high degree of functionality in the searches available to the researcher. For some time there were more ways of searching the texts than there were texts in the collection. In addition to the research benefits of this collection, there is a national benefit of access to general readers. Most of these texts are unavailable in any other form, except in the major research collections and archives of the nation. One of the great strengths of this collection in my view is that lesser known authors are placed alongside the big names in Australian literary history. Passages of text can be easily captured and used in teaching and this will be appropriate at all levels of education, within schools and universities alike.
The research and access benefits of the collection suggest to me an even wider impact and significance. I believe that this collection has the potential to impact upon our national awareness and sense of place in the world. Many of these texts will be re-discovered by Australian readers. The corpus of texts, however, that total body of texts made possible only through their integrated electronic publication, will promote not only a sense of re-discovery, but also a sense of self-discovery. Through this record, Australians will recognise themselves and gain a more mature understanding of what it is to be Australian. This is particularly important at this time leading to the new millennium and to the celebration of the nation’s first 100 years of federation (1901-2001). The country is finding its way through a period of introspection and tentative moves towards republicanism and reconciliation with Aboriginal Australians. The electronic publication of the texts in the SETIS collection will in time have a role to play in this process of national re-evaluation. I am constantly struck by the contemporary relevance of these texts and of how they can speak to us today and inform current debates. I am taken, for example, by Watkin Tench’s bewildered account in 1788 of how the new settlers began to find the diseased and dying bodies of the country’s native inhabitants wherever they travelled around the settlement. Tench could not conceive that the first European settlers had brought disease to the aboriginal peoples of the new land, devastating their populations almost before a shot had been fired in anger between the two cultures, certainly before the devastating conflicts on the frontiers of later settlements. He concluded that the deaths remained a mystery beyond rational explanation, since no such diseases had been reported on the ships travelling from England. I read the liberal, tolerant, European enlightenment-man Lieutenant Tench and I am reminded of who we are and what we have done.

Conclusion.
The electronic Australian literature collection at the University of Sydney Library provides an extraordinarily large range of literature reflecting the experiences of succeeding waves of immigrant settlers meeting and adapting to the challenges of a strange, utterly different geography to the lands of their origin. I believe the collection, as it grows, will assume national and even international significance for its depth and coverage as a chronicle of the emergence in my part of the world of distinctive, new voices through the written text.

The SETIS Australian Literature Database is available at http://setis.library.usyd.edu.au/ozlit