

## The Australian Cooperative Digitisation Project 1840-45 - A Progress Report

---

This article by Ross Coleman and Alan Ventress appeared in the March 1997 issue of *LASIE - Library Automated Systems Information Exchange*, vol. 28, no. 1 (ISSN 0047-3774).

### Abstract

The Australian Cooperative Digitisation Project, 1840-45, also known as Ferguson 1840-45, is a collaborative project between the University of Sydney Library, the State Library of New South Wales (SLNSW), the National Library of Australia (NLA) and Monash University Library, supported by ten other institutional and industry groups. The project has been funded by an Australian Research Council 1996 Infrastructure Grant through the Department of Employment, Education and Training.

- [Background](#)
- [Selection, Preparation, Conservation and Copyright Issues](#)
- [Microfilming Issues](#)
- [Scanning Specification Issues](#)
- [Control Structure, Image Management and Retrieval Issues](#)
- [Optical Character Recognition \(OCR\) Issues](#)
- [Specific Benefits to Researchers](#)
- [Presentation of the Material in a Form that Provides Ease of Browsability and Retrievability](#)
- [Summary](#)
- [Management](#)
- [References](#)
- [Biographical Notes](#)

---

### Background

The project, led by the University of Sydney Library, is based on a preservation and access model, and will provide the digital conversion of significant nineteenth century Australian material to enhance literary and historical research. The full content - text and images - of journals, newspapers, novels and short stories published in Australia from the period 1840-45 will become accessible through the world wide web via the Internet.

The purpose of this collaborative project is to ensure both increased access to, and preservation of, this material through an integrated process of preservation microfilming, digital scanning, network access and selected full-text enhancement. This progress report provides background to the project and an overview of the type and complexity of technical and associated issues faced in the project. Through meeting scholarly research needs the project will be investigating and developing technical benchmarks for similar digital library projects involving the conversion of retrospective print material. It is an opportunity to develop technical and management

experience and skills in such work at both the library and vendor level for the benefit of future projects in Australia.

The need for action to preserve and provide better access to the nineteenth century material listed in Ferguson's Bibliography of Australia had been discussed informally between Ross Coleman and Alan Ventress in association with Stephen Hall of Chadwyck Healey in the late 1980's and early 1990's and through several forums (such as the Towards Federation 2001 conference) as a comprehensive microfilming project with little result.

The impetus for this project came from the first National Preservation Office (NPO) conference in 1994 on the future of preservation microfilming, and particularly from a paper given by Anne Kenney from Cornell University Library on digital conversion for preservation and access purposes (Kenney, 1995).

The possibilities of such a project were discussed between staff from the SLNSW, the NLA and the University of Sydney Library. In subsequent discussions with a focus group of scholars in nineteenth century Australian studies the project received enthusiastic support, with a consensus for the period 1840-45 as the most appropriate for significance of content within a manageable size for the project.

The project will provide access to the published record of life and activity of the period 1840-45, regarded by scholars as seminal in the development of a colonial culture and an emerging sense of an Australian identity in the colonies. This period, following the end of convict transportation to New South Wales and preceding the influx of the gold-rush migrations, heralded the introduction of representative government in 1842 and witnessed the early years of mass free migration to the Australian colonies. It was a period marked by an upsurge in local publication in both the older and newer centres of settlement to meet the demands of the increased population and the commercial opportunities emerging in the colonies.

This project will concentrate on two key aspects of publication of the period. The first, covers the content of journals and newspapers that began publication in the period. Seventy-five titles are listed in Ferguson, and additional non-Ferguson titles have been identified, and these will be scanned up to 1855 (only six titles were still being published after 1855). These titles represent a largely untapped primary resource for historical and literary study. The second aspect of the project will provide the full text of the novels and short stories published from 1840-45, significant because they were among the first to concentrate on descriptions of bush life which was to be so dominant in later fiction. These scarce, culturally significant resources, mostly held in either the Mitchell Library of the SLNSW or the NLA, will be made directly accessible to researchers, libraries, schools and individuals in Australia and overseas. Images of all material will be accessible via the Internet through the NLA in the later half of 1996, for browsing and down loading if required. Material down loaded can be further enhanced by users to suit their particular needs. The fiction of the period will be further converted through optical character recognition (OCR) software by Monash University Library to be accessible and fully searchable in ASCII format as well as viewing page images. Technical and user specifications are being developed and the production work of filming and digital conversion will be

contracted to commercial vendors. Other products of the project to be investigated include CD-ROM on demand and print reproduction.

In this project we acknowledge and draw upon the experience of similar international projects. The key coordinators of two of the major American projects, Anne Kenney at Cornell (Cornell Digital Library) and Paul Conway at Yale (Open Book Project) are acting as international advisers to the project. Other projects taking place that have assisted or will assist in our work include the Library of Congress digital project, the Making of America (Cornell and Michigan Universities) and the UK Elib sponsored project, the Internet Library of Early Journals. In turn we hope our experiences will assist other projects in this area of digital library development. In investigating technical benchmarks for the large scale digital conversion of retrospective print collections, production specifications will be developed for carrying out this type of work through existing commercial vendors in the field. This will involve testing their capacity to handle work on a reasonably large scale to standards that will enable high quality replication of the originals. This also involves the establishment of quality control standards for both microfilming and scanning, and presentation of the digital files in an appropriate file structure. Each stage of the project must be undertaken to ensure that the final product is useful to the user as a resource for study and research. A great deal of effort has gone into the early stages of preparation, conceptual planning and testing to ensure this goal is reached. Each stage presents its own set of issues to be considered, and many more will arise as we work through the project.

### **Selection, Preparation, Conservation and Copyright Issues**

This has involved the identification and checking of the original material and has proven to be rather complex, much more so than anticipated. Identifying complete sets, or filling gaps when parts of titles may be held by different institutions has been part of this process and has required the input of staff with very highly developed bibliographic skills. It has also required the selection of the best copy for the project. This should be not only the best copy physically (for which criteria must be observed), but where practicable with significant contemporary annotations that may add scholarly value and interest. A good deal of the material in the time period of this project has already been microfilmed. Some consideration has to be given to the utility of these existing microfilm sets as a source for scanning, particularly in regard to quality and page orientation on the film (the digital image must appear in "portrait" style with the image appearing on the screen as straight up and down). Older microfilm will in many cases be replaced by technically superior modern film to ensure the highest fidelity is achieved during the digitisation process. While the print material itself is out of copyright, issues such as rights over the master microfilm, and rights over the final digital images need to be considered. Where necessary some initial conservation work may need to be done on the original material. Fundamental to the project is the maintenance of the integrity of the original material. The material is scarce and of cultural significance, so no actions are taken that may damage the originals. Nevertheless in certain circumstances the disbinding of books, journals and newspapers may be necessary to allow the microfilming and subsequent digitisation of the material to the highest possible standard without "gutter shadow".

### **Microfilming Issues**

This has involved a thorough investigation of existing preservation microfilming standards to establish if they are appropriate for the project. These standards have been accepted, with the added specification that the material is in a layout that will assist production scanning. The standard and quality of the microfilm image and the preparation for microfilming is important as this will affect the efficiency of the production scanning process. Attention to the image at the microfilming stage will minimise the need for intervention during the scanning process.

### **Scanning Specification Issues**

The development and writing of appropriate specifications for this type of project in Australia are based on local experience at the National Library of Australia and State Library of New South Wales and the tremendous amount of work done in the United States, particularly at Cornell. Alan Howell, the convenor of the Technical Advisory Group, also attended a workshop at Cornell in 1995 which has provided much valuable input. (Howell, 1996). Some of the issues have been the investigation and testing of options and appropriateness of various image resolutions (bitonal or grayscale) as well as image depth (at 200, 400 or 600 dots per inch [dpi]) to provide the best digital replication of the image in the most efficient file size that will allow fast access across the World Wide Web. Issues of file compression, consideration of standard digital file compressions (TIFF CCITT group 4, JPEG) or new developments in compression, such as fractal (FIF fractal image format) compression. All of these issues have proved to be critical parts of the image management and retrieval system which will effect the final database design and the user interface.

### **Control Structure, Image Management and Retrieval Issues**

This area of development has been assigned as the prime responsibility of the National Library and has required the investigation and development of a system architecture, the provision of a document control structure and a management and retrieval system that will allow efficient browsing and retrieval of images. This underlies the success of the project and a great deal of effort has been given to planning, testing and consultation prior to the commencement of production work. The development of directory structures for vendors to present the scanned files to allow efficient loading of the images into the document control structure is also critical to the success of the project as is consideration of the need for off the shelf, preferably "public domain" viewer software that will enable efficient browsing, zooming, down loading etc. File types, compression and file size are all issues which have a bearing on the final utility of the planned product.

### **Optical Character Recognition (OCR) Issues**

This segment of the project has been delegated to Monash University and requires the selection and use of appropriate OCR software to scan in and read the TIFF files of the four novels which have already been created for this phase of the project . The need for proof checking and correction of OCR'ed files will be required, as will programming to enable the ASCII text of the page and the image to sit side by side.

### **Specific Benefits to Researchers**

Network access to relatively scarce and physically vulnerable material of scholarly significance to researchers, libraries, schools and individuals in Australia and overseas will be provided through the project, this will radically and significantly extend support for research and study of Australian studies worldwide and will provide longer term access to this material by preservation of the content through microfilming. Network access will also reduce use of the original material ensuring the preservation of its physical integrity, at least for a little longer. Provision of this material in digital form will also give users the opportunity to download and enhance (eg by OCR conversion to ASCII and further enhancement by mark-up languages) or manipulate the material as best suits their particular needs - be they historical, literary, linguistic, genealogical or teaching.

### **Presentation of the Material in a Form that Provides Ease of Browsability and Retrievability**

The project will enable the delivery of this material by other means such as CD-ROM or print reproduction and will create digital content of heritage material for further industry and multimedia production.

### **Summary**

This brief description of many of the issues currently being addressed in the project only gives some indication to the complexity of the work involved. Throughout this project we will address many of the major technical and management issues of large-scale digital conversion, and develop new partnerships in collaboration between major institutional libraries, and between librarians, researchers, microfilming companies and the nascent Australian digitisation industry. The project will assist in the development of technical specifications and skills based on international best practice, and will present a new level of resource-sharing between libraries.

### **Management**

The Australian Cooperative Digitisation Project 1840-45 is chaired by Professor Elizabeth Webby, Professor of Australian Literature at the University of Sydney. The project Steering Committee is Ross Coleman (University of Sydney, Convenor and Project Manager), Robert Stafford (Monash University), Alan Ventress (SLNSW) and Colin Webb (NLA). The Technical Committee is Alan Howell (SLNSW), Diana Dack (NLA), Jan Lyall (NPO), Anne Kenney (Cornell) and Paul Conway (Yale). Site Coordinators for the project are Andrew Heath & Rebecca Thomas (SLNSW), Jutta Hosel (NLA) and Julie Price (University of Sydney).

### **References**

Howell, Alan, 1996. "Digital imaging technology for preservation and access: a Cornell University Library workshop". *LASIE* 27 (1): 26-41.

Kenney, Anne R., 1995. "From analog to digital: preservation reformatting in a changing world", in *Preservation microfilming: does it have a future?*, Proceedings of the First National Conference of the National Preservation Office, at the State Library of South Australia, 4-6 May 1994. Canberra, National Library of Australia, 90-106.

## **Biographical Notes**

**Ross Coleman** is currently employed as Collection Management Librarian at the University of Sydney Library where he has been since 1991. Prior to that he worked at Deakin University Library, Nepean CAE Library, The British Library and Macquarie University.

**Alan Ventress**, was appointed to the position of Mitchell Librarian in 1993. Between 1987 and 1993 he was Manager Resource Networking & Collection Development at the State Library of NSW and was editor in chief of the Library's collection development policy. At present he is working on various projects to digitise the Australian Research collections including the Banks Papers on Internet project and the Australian Cooperative Digitisation Project.

Ross Coleman, Collection Management Librarian  
University of Sydney Library  
University of Sydney NSW 2006  
Telephone: (02) 9351 3352  
Facsimile: (02) 9351 7305  
E-mail: [collm@extro.ucc.su.oz.au](mailto:collm@extro.ucc.su.oz.au)

Alan Ventress, Mitchell Librarian  
State Library of New South Wales  
Macquarie Street  
Sydney NSW 2000  
Telephone: (02) 92301466  
Facsimile: (02) 92351687  
E-mail: [aventres@ilanet.sl.nsw.gov.au](mailto:aventres@ilanet.sl.nsw.gov.au)