# Digitising the record of a colonial culture - Ferguson 1840-45

*[Published in : Moving Theory into Practice, digital imaging for libraries and archives. Edited by Anne Kenney and Oya Rieger.*
*Mountain View, CA, Reserarch Libraries Group 2000]*

There is a certain circularity about content in any digital project. Initially, at selection, content is the, subject, substance and value of the endeavour. As the project moves into production, content, as images, becomes a purely technical, objective, issue dominated by concerns about physical characertistics - the quality of imaging, the determination and delivery of file sizes become the all absorbing obession. After the 'digitisation' phase, and access is provided, the value of the content as subject returns, tempered by the quality and standards of the digitisation phase. In our practice we are closing this circle and understand the process through the delivery of content.

This sidebar will reflect on the implications of the decision to digitise - comprehensively - the contemporary publications of a period of particular significance for Australian political, historical and literary study. That selection decision (and the context of that decision) determined both the importance and character of the project, and the nature of many of the subsequent technical processes. The decision required a collaborative approach, and (at the time) use of a film/scan model. While a major challlenge was the diversity of formats and condition of the original material, the most difficult process for all involved was the practical application of a set of specfications and standards that had been developed, based largely on documentation from other projects. This sidebar describes a steep learning curve where theory and enthusiasm met the practicalities of application and demands for quality output.

The period 1840-45 was a seminal period in Australian history, recording and reflecting the emergence of an independant colonial character. The 75 periodical publications of this period, as listed in Ferguson's *Bibliography of Australia*, the most comprehensive listing of publications pertaining to Australia, formed the primary subject of the collaborative Australian Cooperative Digitisation Project, 1840-45. The collaboration - between three of the largest research lbraries in Australia, the National Library of Australia (NLA), the State Library of New South Wales (SLNSW), and, the project lead, the University of Sydney Library - was funded in 1996 by an Australian Research Council (ARC) Research Infrastructure grant. This grant program, a competitive University-based program to fund anything from science facilities to resource projects, was the only likely source to fund a developmental project of this scale.

The interest in undertaking some form of major digitisation project in Australia followed the First National Preservation Conference - 'Preservation Microfilming, does it have a future?' - held in 1994. For some time there had been discussion about systematically microfilming the titles listed in Ferguson, but a paper by Anne Kenney on preservation microfliming and digital access models provided the impetus for this particular approach.

The access benefits of digitisation also caught the imagination of many academics who used, as primary sources, scarce nineteenth century material held only in the collections of the NLA and SLNSW.   The proposal for a preservation/access project was canvassed with a small group of leading academics working in Australian literature and history.   Their response was enthusiatic, and their immediate consensus was to focus on the years 1840-45 as a key period in the development of Australian colonial culture. Of more consequence to the project, however,  was their insistence on taking a comprehensive approach to the period - the value of the project, they argued, was in the full record of the period.  This approach, with wide academic support  from a number of large and small insitutions gave the project sufficient academic support to gain ARC funding.  The proposal to digitise also gave the project the innovative edge.  So the scholarly significance of this comprehensive approach also assisted to secure funding.

This decision to take a comprehensive approach immediately highlighted one of the primary benefits of digitisation - the merging of distributed collections into one complete virtual collection.  Such an endeavour required an effective collaborative organisation. The emphasis on comprehensivenes required the checking of holdings of several libraries identify and fill gaps in the collection.   The detail and effort involved in checking holdings down to the issue level, the identification of 'best copy' for microfilming, checking if titles were already microfilmed and whether the film was suitable for digitisating was our first serious underestimation of the level of work and time this project would demand from us.

The adoption of the 'film then scan' method was both content driven and pragmatic. The importance of maintaining the physical integrity of the fragile and scarce original material was a primary consideration that determined the method.   Microfilming was considered (at the time) the most effective and non-damaging means of reformatting the original material into a form that could be digitisted in a production process. The microform masters provided the long-term, stable, preservation format. Pragmatically, there was substantial institutional and vendor expertise in Australia with microfilming, and liltte with the kind of digitisation proposed in the project, so this seemed the most appropriate and familiar technical method.

The determination of method was stengthened by the decision to contract the production work, rather than carry it out inhouse.  The need to foster industry skills by developing vendor experience in digitising heritage material was a fundamental element of the project. This was envisaged as a long-term benefit of the initiative. This decision (given the level of inexperience and naivette among both the partners and vendors) provided many months of agnst and delays while both parties tested and retested  technical specifications, worked to resolve production complications, and reviewed the quality of output (again the effort and time involved was unanticipated).

The preparation, reviewing and application of technical specifications and standards became our most serious concern, and this was complicated by the diversity of material being handled in the project.  In the RFP process we documented, and liberally adopted in theory, specifications based on work at Cornell, Yale and LC. We initially tested these specfications in a test phase to provide some practical experience and a basis for expected image quality from the originals.   Imaging

standards were determined as 400dpi bitonal. The awarding of the contract to a partnership of a leading microflim vendor and a small multimedia company with some experince in dealing with heritage material shifted the project from test to production. Microfilm specifications were based on RLG recommendations, and issues addressed ranged from standard reduction ratios, targeting, blips, layout, masking and process routines to facilitate high speed digital conversion using a Sunrise microflim scanner. During digitisation procedures were established and reviewed for for cropping, de-skewing, cleaning of images; standards were established and reviewed for acceptable image quality; processing and naming protocols were established and tif header and metadata elements specified. Image quality standards, initially based on testphase output produced by a Screenscan scanner needed to be reviewed to allow for the production scale output of the Sunrise scanner. We experienced classic comunication problems on both sides (such as differing interpretation of specific instructions) and our production expectations were too high (as they were based mostly on the images produced in the 'handcrafted' test phase, compared to the exigencies of the production process) - it took several months for agreement on acceptable image outputs. It was necesary for us to stand back to understand the process from the vendors view, but still to insist on basic quality standards. The learning curve was a steep, and often frustrating one, for all parties.

The most significant factor in digital conversion and production was the nature of the original material. Having taken a comprehensive view of the content we were committed to handling all the physical variations of the publications - not only in size, from quarto to broadsheet - but also the variety within any one title - from foxing and discolouration, to the use of varying fonts and point sizes on the one page. This diversity of sizes and variation of condition presented challenges for image quality, file size and delivery mechanisms. Some issues are still to be resolved, so, while we have scanned broadsheet newspapers, effective web delivery (other than TIF delivery of individual pages) has not been fully tested through PDF, our delivery format.

Use of PDF for delivery was determined after consultation with user advisory groups. Examples of the images in TIF, GIF and PDF formats were provided for comment, prior to the decision to use PDF. Though image files are not the most compatible with PDF, particularly because of file size problems (PDF is more suited to text documents than images) - it still provides a most effective delivery package. Though using PDF for delivery we will retain the production TIF files on CD as digital masters.

At the end of the production stage we have completed 65 of the periodical titles - the other 10 titles either could not be located or were only available in microform of a quality that was not considered of convertible standard (though that standard has shifted as vendor scanning expertise has improved). To add access value to the journals we will OCR all internal indexes or contents lists that exist for these titles, and undertake to OCR the fulltext of several titles to investigate the technical issues and processes in providing an uncontrolled keyword search facility.

As the first major project in Australia to digitise on a large scale nineteenth century material we began from perhaps from an overly ambitious base, imaging a comprehensive record of significant historical material, whilst trying to establish a

contract production process. While we have passed through the production phase, we are still in process of preparing the material for delivery. And while this path has been - like many projects - cast with difficulties and frustrations - the benefits are now becoming apparent. The material available so far is being used for study and research around Australia, the vendors are being offered work from a growing number of clients, the collaboration is continuing, and each partner has established in different ways ongoing digital units or services pursuing individual and collaborative projects and programs. Throughout the Ferguson project, and the other projects we are pursuing, it is the value of the content that underpins and substantiates the endeavour.