Digital conversion of Nineteenth century publications - production management in the Australian Cooperative Digitisation Project 1840-45

Ross Coleman, University of Sydney Library r.coleman@library.usyd.edu.au
Colin Webb, Director, Preservation, National Library of Australia cwebb@nla.gov.au

Abstract

This article will discuss and evaluate the management and production issues of the Australian Cooperative Digitisation Project 1840-45 - a collaborative project funded by the Australian Research Council. The completion of this digital library project allows the authors to reflect on the technical issues and the interrelationships of the content, project organisation, the production model with its access and preservation goals, and issues of quality control, in relation to the future viability of such projects.

> *It may in truth be said, that in no country, and at no period since the invention of printing, has there appeared a greater necessity for a periodical conducted with spirit and principle, in the popular cause, than there does appear to exist in the colony of New South Wales at this moment. The only journals of character now existing upon any basis that offers to be permanent, boast of their "moderate conservatism"; a term which according to our interpretation means just as much oppression of the many by the few as the spirit of the age will bear.....unless the people are fully represented in the periodical press, as well as in the councils of the country, their rights will in the conflict of factions and interests be greatly endangered"*
>
> The Weekly Register of politics, facts and general literature, vol1 no 1 July 29 1843.

The period 1840-45 was a seminal period in the development of an Australian colonial culture. This period, following the end of convict transportation and preceding the influx of the gold-rushes, was characterised by exploration and expansion, conflict, commercial growth, political agitation and a surge in local publication reflecting the issues and concerns of the time. Journals, such as the Weekly Register, led and engaged in the political and social debate of the time, and remain today as the voice and contemporary record of the period.

The Australian Cooperative Digitisation Project, 1840-45 (ACDP) was funded, through an Australian Research Council grant, to both digitise this contemporary record for access and ensure its long term preservation. The project has been a collaborative initiative between the University of Sydney Library, the State Library of New South Wales (SLNSW), and the National Library of Australia (NLA)

Fundamental to the success of this project was the need to establish practical and implementable standards for large-scale digital conversion, in the context of the hybrid (microfilming and imaging) production model adopted using external vendors. The project - following the access and preservation initiatives developed in the US -

has been described in a number of earlier articles (these can be found at the project site at http://www.nla.gov.au/acdp/), and we do not intend to revisit these descriptions in detail.

This article will address and evaluate the management and production issues of what has been a complex developmental digital library project.. This complexity can best be characterised by the interrelationships of the nature of the content, the project organisation, access and preservation goals, production issues and management, and image quality control.

Nature of the content

The content of the project can be understood in three ways -:in terms of the intellectual significance of the content, the physical nature of the original material, and the nature of digital content itself.

Intellectual significance

The primary purpose of the project is to enhance literary and historical research on nineteenth-century Australia by providing improved access to, and preservation of, scarce primary material confined to a few major library collections. The period 1840-45 was selected by an advisory group of academics as significant for recording, for the first time, the emergence of a particularly Australian political, social and literary character. Their advice, to treat the period comprehensively and to cover all the journals that began publication in the period as well as the novels, determined the character of the project and dictated many of the subsequent technical processes. The bibliographical source for material was Ferguson's *Bibliography of Australia* - the most comprehensive bibliography of nineteenth century Australian material, from which 75 journals and four novels were initially identified as the subject of the project.

Nature of the original material

Having taken a comprehensive view of the content we were committed to handling all the physical variations of these 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.

The selection and assessment process - where each title was verified, located, identified as ‘best copy’, gaps identified and remedied by access to other copies, and assembled for microfilming presented a different array of challenges requiring detailed bibliographic skills and untold patience. The documentation of the assessment process on a title-by-title basis became a critical management tool for monitoring and controlling workflow.

As the project developed, we learnt just how important this final pre-filming assessment was. It was impossible to achieve satisfactory results by treating all materials the same way. Material had to be assessed more closely than we had

expected to identify the likely challenges and quirks of each title. This was similar to normal microfilming practice, but with the discipline imposed by later scanning of the microfilm, this pre-film sorting and description of the material needs to be much more rigorous.

From the outset we were committed to ensuring the physical integrity and conservation of the original material throughout the process. This concern was met in part by the decision in our production model to microfilm first as we were confident, from experience, in the ability of microfilm agencies to handle fragile heritage material in an appropriate manner

Nature of the digital content

The technical aspects of the digital images (file size issues, delivery considerations, quality issues) are detailed in the discussion of production issues, below. The pertinent point here is to appreciate that through digital conversion and reformatting the value and potential of this content has also be transformed in fundamental ways.

The decision to take a comprehensive approach immediately highlighted one of the primary benefits of the digitisation of this material - the merging of distributed physical collections into one complete virtual collection. Such an endeavour required both a national viewpoint and an effective collaborative organisation

Project organisation and management

ACDP was intrinsically a collaborative project. This presented great benefits as well as challenges for the project partners and vendors involved in the collaboration. The challenges included managing the flow of information, ensuring inclusive decision-making, maintaining adequate documentation, monitoring progress, and adjusting procedures. On the positive side, the collaboration provided a wide pool of expertise, experience, motivation, and diversity of ideas and styles that all added greatly to the project, helping us deal effectively with most issues. These benefits extended well beyond ACDP to other projects.

The vendors (chosen through selective tender) themselves needed to set up the required infrastructure (investing in expensive imported hardware, and undertaking considerable programming and software development) to provide the level of throughput and quality required by the project.. The working relationship between the project partners and the vendors was a key part of the project, though sometimes rocky as we tested and reviewed outputs against our specifications (we went through 10 testing and feedback stages - see Appendix 1). The decision to outsource by contract was fundamental to the project - we wanted to use this opportunity to assist the development of industry skills among contractors in the field, and to facilitate the investment and foster the expertise for use in similar projects in the future. This has been one of the many successes of the project.

Appendix 2 lists Facts and Stats of ACDP

Project Management issues

The first production issue was to recognise the complexity of this sequence of steps and objectives, and the need to manage them closely and well. Often the issues that arose in managing workflows and communications seemed the most critical of all.

The exploratory nature of the project, the interdependence of processes, the number of players involved, and the distance between them – all meant there was potential for many risks including confusion, delays, loss of accountability, and important things falling between the cracks. The project management team recognised many of these risks and sought to manage them proactively with a variety of tools and responses, including:

- appointment of a dedicated project coordinator (located at SLNSW)
- preparation of detailed specifications
- establishing a project management structure consisting of a high level Project Management Committee, a more operationally-focused Project Steering Committee, and a Technical Advisory Group
- allocation of operational responsibilities between partners
- agreeing on communication guidelines and mechanisms, including meetings, and regular email and phone contact between meetings
- negotiating communication mechanisms with vendors, including regular meetings, reports, and other feedback opportunities
- negotiating timelines for expected deliverables
- developing tables of titles and what had to happen to them
- batching of work

Project management inputs turned out to be a very large – and largely hidden – investment. We came to recognise the reality of the often-repeated assertion that the organisational and management aspects of digitisation projects – and especially complex collaborative ones – may be more challenging than the technical issues.

Access issues

This project was devised and designed to serve preservation objectives while greatly improving the accessibility of material that had previously been difficult to access.

Internet access was the rationale behind the project, so the main access issues were to produce file in formats that could be effectively delivered to users over the Internet, and to develop a user interface that would be easy to use.

Formats and files sizes

Bandwidth is a major problem for delivering large, high-resolution images over the Internet, and it was inevitable that this would be a challenge for ACDP. Legible images of difficult, small print text necessarily produce large file sizes. And, as the dimensions of the original material increase, file sizes increase geometrically.

We also found that, while images of photographs and painting can remain worthwhile for reference purposes even at low resolutions, text images are useless if they are illegible. Finding ways to deliver legible images at acceptable speeds became a key focus. These issues needed to be resolved at the earliest possible stage so that production work would support, rather than complicate, the eventual storage and access systems. Development of the system architecture, document control structure, image management and retrieval system, and user interface - undertaken by NLA - ensured a consistent approach to services delivered from its website.

Following tests of a range of delivery formats with the project's prime target audience, Adobe's Portable Document Format (PDF) was chosen as the online delivery format for the compressed bitonal TIFF images. A number of images (between 1 and 10) are embedded in each PDF file, providing multi-page files for easier moving between pages. PDF also offered advantages for zooming and for printing, important features given that whatever the format, most pages cannot be viewed on screen as a whole at a legible magnification. The need to download an Adobe viewer was acknowledged as a possible barrier to access, but due to the increasing likelihood that users would already have suitable viewers installed, this was not seen as an impediment.

Possibilities for searching text

The great majority of images produced are bit mapped images of text, rather than text that is searchable. Because most of the original serials lack indexes, finding specific content is difficult. This made the project partners interested in the possibilities and potential of using optical character recognition (OCR) software. However, the characteristics of the original material – small and variable typefaces, poor print quality, and complex arrangement of print on the page - meant that available OCR software was ineffective. Interest in using OCR for the serials remains, and further trials are now being conducted with more satisfactory results.

User interface and access points

The user interface allows browsing of titles, and searching by title, year of first publication, and broad subject. Users can also access detailed information on how to use the site, notes on the digitisation of each title, and background papers and reports on the project.

Following NLA practice, both MARC bibliographic records and Dublin Core resource discovery metadata are used as points of access from outside the project website. There are direct links to titles from Kinetica; bibliographic information is also included with images in the interface. For each title reproduced by ACDP, catalogue records

are created for the original hardcopy version, for the microfilm version, and for the digital file.

Preservation issues

Preservation considerations were manifested in four ways: attention to protecting the originals from damage during the process, aiming for image quality that would minimise the need for users to access the originals, use of microfilm to provide a long lasting analogue preservation copy, and attention to the archiving of the digital files.

Care of originals

Although most material was more than 150 years old, it was generally in reasonable condition. Much had been printed on rag paper, and showed little sign of brittleness. There were occasional legibility difficulties, due to poor contrast between faint print and darkened (but not brittle) pages. Apart from some minor repairs to tears, very little conservation work was needed to deal with physical damage. This was in line with developing best practice: tailoring microfilm-preparation conservation work to the minimum required so material can be safely and effectively filmed.

Image quality and preservation of originals

The use of the source material by researchers was relatively low - indeed the academic advisory group commented that this material was under-utilised because of difficulty of access. However the preservation requirements would not be realised if the digital images simply stimulated users to increase their demands on the originals, so image quality in the microfilm and digital versions had to be sufficiently high to meet most research needs.

Use of microfilm

Use of microfilm as an intermediate source for digital imaging, a master for re-scanning, and a long-term stable storage format was a key feature of this project.

At the start of the project, we were convinced of the need to produce microfilm as a reliable preservation copy. We are no longer convinced that digital copies cannot provide a suitable preservation medium, although it remains to be proven over extended periods of time that digital images will be cost-effective preservation masters..

One should also not be too starry-eyed about microfilm as a preservation master: it is expensive to store properly, requires some vigilance to detect deterioration, and in most cases provides a less-than-perfect replica of the source. However, it remains true that poorly maintained microfilm is more likely to yield at least some of its information content than poorly maintained digital data.

Archiving the digital files

The preservation aspects of the project did not depend primarily on long-term maintenance of the digital files, but given the very large investment in them and the unlikelihood of ever assembling material to repeat the exercise, the project partners paid some attention to that outcome.

There was a concern to make the digital images worth preserving. This largely came down to a question of capturing all significant detail. From tests it was established that 400 dots per inch bitonal images would substantially achieve that. Bitonal scanning cannot be accepted as a totally satisfactory replica, and the realities of scanning make it is difficult to define the smallest significant detail that will be captured or not captured. However, the project partners agreed that this approach would produce images that would be worth preserving.

There was also concern to use a file format with a reasonable chance of being preserved. At the time of the project, this was clearly TIFF, (Tagged Image File Format), a non-proprietary standard which should be relatively straightforward to migrate, and with sufficiently widespread use for there to be a critical weight of demand for migration tools. The format also offered good lossless compression, and the ability to re-use the images in a number of derivative formats. It was (and is) widely accepted as the closest thing to a preservation format for digital images.

To provide the kind of active archival management that will be needed requires systems, procedures and commitments. As part of the project, NLA has accepted responsibility for maintaining both microfilm and digital versions of the ACDP output.

Production issues

The basic production issue was how to achieve the goals of access and preservation objectives cost-effectively.

To gain some sense and appreciate the complex processes involved in production, one need only consider the exhaustive steps that project managers and vendors needed to manage, once specifications and contracts were in place:

- locate material from a range of sources amongst the participating partners and elsewhere, negotiating release of additional material for filming if necessary
- choose copies for inclusion from multiple copies held in various collections and in various conditions
- check and prepare material including any necessary conservation work
- assemble material at a central point and provide specific item and batch notes
- transfer material to the vendor
- microfilm material to specifications and produce duplicates for checking by both vendor and project manager
- vendor quality control checking and documentation
- transfer duplicate film to the project manager along with original material and documentation if needed to support quality assessment of the film
- quality control checking of microfilm by the project manager (at SLNSW)
- accept or reject the film and advise the vendor, returning material for re-work if necessary
- re-work of microfilming, duplication, and quality assessment if necessary

- duplicate the film to produce first-generation preservation master, second-generation duplicating master, and third-generation service copies
- assess suitability of the film (both pre-existing film and that prepared for the project) for scanning
- scan the second generation film if suitable
- enhance the images if necessary
- vendor quality control checking of digital images, and documentation
- name the files, apply specified metadata, and save to CD-R
- transfer the CD-Rs (and if necessary for quality control checking, the microfilm and the original material) to NLA
- quality control checking of digital files by NLA
- accept or reject the digital files and advise the vendor, returning material for re-work if necessary
- re-work of digital scans, image enhancement, metadata, file names, file saves or documentation if necessary
- repatriation of original materials to owners and supply of microfilm service copies
- convert the digital TIFF files to access format for online delivery to users
- mount the access files on server
- possibly apply optical character recognition (OCR) to selected titles, assess, edit and mount on server
- archive and maintain digital copies in all formats selected for long-term maintenance, as well as microfilm preservation and duplication masters (and originals).

Project Management

The development, and review of specification, a process of testing and feedback to vendors, and the requirements for and coordination of dedicated staff resources were among the realities underlying project management

The project invested heavily in developing detailed specifications to cover all production processes, with an emphasis on achieving suitable image quality..

For all the work put into specifications, vendors frequently complained of difficulties in complying with them to achieve the results we wanted. (This may have been because the specifications were wrong, or because they had been misunderstood, or because they demanded solutions the vendor was unable or unwilling to use, or the project could not afford)

By the end of the project we accepted that our initial specifications may have been too focused on how results were to be achieved. We concluded that for large projects, it would be better - pragmatically - to have an overall framework of objectives and ‘bottom lines’ negotiated with vendors, supplemented by detailed specifications for small batches of material.

Testing and feedback

One of the project management tools we valued most was a feedback approach – tests, reports, communication of experiences between vendors and project

managers (“this worked, this seemed not to work”), and subsequent decisions about the need to adjust procedures. Such an approach was intrinsic to the project management philosophy of ACDP, and is critical in such a developmental project.

It is possible to see this ‘feedback’ approach at work throughout the project: the whole production experience can be seen as a series of tests, interspersed with production phases In total we went through 10 testing phases (see Appendix 1), and while this may seem excessive it was both appropriate and necessary, particularly given the developmental nature of the project.

Decisions on how to respond to feedback were mostly pragmatic. On some issues we adjusted procedures because we could see the necessity to do so; on some we adjusted because we accepted it would be impossible to achieve what we wanted in the circumstances; on some issues we decided not to adjust because we were still convinced we were right; and on some we decided not to adjust because the project had been set up to test certain things and we felt it was too late to abandon them. Our ultimate conclusion is that it is not possible to eliminate these tensions – it is frustrating to delay production for feedback to be analysed, but the feedback is necessary all the same – at best one can hope to strike a good balance. Likewise, it is very unlikely that project specifications will serve without some adjustments and compromises – again, the best one can hope for is to identify the really crucial requirements and try to negotiate robustly, honestly and reasonably to achieve them.

Dedicated staff resources

The allocation of dedicated staff resources to the project was tremendously important, as was the commitment, initiative and good humour of all the people involved. However, many of the issues were just too big for the part-time resources allocated, and simply required more input, more sorting out, and more hands-on management than was available. We concluded that life would have been much easier with a full-time coordinator in each of the main operationally-involved libraries (NLA and SLNSW), rather than a part-time coordinator in one institution supported by less-than-part-time staff trying to fit the project into other demands.

We also learned something about the effects of such a large project on the people involved. The ARC grant for the project was approved in late 1995; the final production phase Request For Tender was issued in February 1997; digital capture was completed in March 1999; file mounting was completed in early 2000. This timetable far exceeded anyone’s expectations, and imposed its own difficulties as a number of people moved on to other things before the project was completed. (Of approximately 35 people with significant involvement in the project, only the three members of the Steering Committee maintained their active role from start to finish.) Maintaining continuity through such a lengthy and complex exercise required a strong project plan and documentation and continuing institutional support. Smaller projects that could be completed within a shorter timeframe would have many advantages in comparison.

Image quality production issues

Many production issues centred on image quality and how to achieve it. In particular we needed to explore how high quality microfilm images would convert to ensure high quality digital images.
Management of image quality in production basically occurred on two fronts: the production processes, and quality control.

Microfilming, scanning, and image enhancement

Most production challenges arose from the interaction of microfilming, scanning and image enhancement, and the often conflicting demands of image quality on the one hand, and production throughput on the other. The thrust of technology and project management has been to find a workable compromise that maximises both image quality and throughput.

Two of the project partners and the project vendor were able to draw on extensive experience with preservation microfilming. Apart from a few errors, the microfilming processes of the project went well, producing high quality film that would meet all relevant standards.

However, automated scanning processes demonstrated that more consistent image placement, skew, and masking of edge and gutter shadows were needed. More consistent image contrast is also an issue, though contrast variations are intrinsic to the material being filmed and largely beyond the control of the camera operator.

We eventually asked the vendors to do test scans to work out the best means of maximising legibility and throughput.

Detecting the edges of images was a key requirement for the film scanner and for post-scanning processing software. Skew and gutter shadows were found to complicate this, especially if text came close to the edges of the page or close to shadows that sometimes appear in microfilm images. Both image quality and throughput were affected by this.

A suite of software was used for image enhancement. These tools fit best into a production model where they can be automatically applied to all images with a minimum of human decisions. This brought its own problems, as it was hard to determine at production levels just how much enhancement was a good thing, and how much was too much. De-speckling, for example, can actually reduce legibility by eliminating parts of letters.

We reached a threshold point for image enhancement dilemmas on a number of titles that included very faint, broken-up print, or very discoloured paper, or very obtrusive bleed-through, occurring inconsistently through a volume, and sometimes even across a single page. We also found that automatic de-skew and cropping could remove text that encroached on margins on badly skewed pages. Customising enhancements for parts of volumes or even parts of pages, drastically reduces output levels. If this is not built into the pricing structure and budgets for the project, it tends to create more pressing deadlines elsewhere in the project, so the ultimate effect is felt far beyond the individual images that may have been initially involved.

In some cases we were led to ask how much difficulty was too much for the material to be worth scanning. A small amount of material was rejected from the project on this basis. At the individual page level, however, we sometimes found we had to accept quite marginal legibility when all reasonable options had been exhausted. In these cases we have included a message referring to the difficulties of the original.

Use of greyscale at lower resolution levels was one of the options tested. This gave a more legible result than high-resolution bitonals for some poor contrast material, but produced no improvement at all in some cases, and came at the cost of very large file sizes, so it was generally not used.

Overall, the project pushed the microfilming end as far as it could go for some material; to make improvements the only remaining leverage rested with scanning and image enhancement. There is some give and take between filming, scanning and post-scan enhancements, demanding conscious decisions about what manipulations and enhancements will work most cost-effectively at what stage.

These trade-offs, and the complications that one process can cause for later steps, give a powerful advantage to arrangements which make the same firm, and possibly the same operators, responsible for the entire chain of processes.

Assessing existing microfilm for scanning

While the project originally sought to make maximum use of pre-existing microfilm, very little suitable film was found to exist and most material was scanned from microfilm created specially for the project. However, assessment of existing microfilm for scanning purposes remains a key concern of the project partners.

How does one decide if existing film is suitable for scanning? While it is not very helpful, the answer is really that it depends on what image quality is required at the end. It requires decisions about the minimum level of legibility and aesthetics that will be acceptable, and the resources one is willing and able to make available to achieve that minimum or something better. It also requires a deep level of interaction between project managers and vendors to articulate, test, refine and establish what is achievable. Only then can one decide whether “the best we can get” will be worth the cost and effort.

Quality control and image quality

In line with our “feedback” philosophy, quality control was a key production issue. Vendors were required to thoroughly check all deliverables, and the project managers also set up quality checking arrangements for microfilm (at SLNSW) and for digital images and files (at NLA).

There has been widespread discussion of tools for quality checking of digital images. While there is undoubted value in improving these tools, we were convinced that the main quality checking requirements could be met pragmatically with simple tools and approaches – the main requirement is that the images be assessed. Our assessment fell into two categories:

- use of software to check file formats, file names and header information, resolution, compression type, and number of files. This was applied to 100% of files.

- subjective visual inspection – looking at the images on-screen and in print-outs, for legibility, polarity, skew, orientation, image cleanness, and matching file names.

Although we were aware of other approaches that attempt to measure image quality objectively, we were satisfied that our subjective inspections told us what we needed to know, and provided an adequate and agreed basis for negotiation with our vendors. This was done on a basis of random sampling at various rates, although for many titles we did check all images.

The use of technical targets, familiar from preservation microfilming practice, proved to be problematical. We decided to incorporate some targets that could be compared across film and digital scans, but we recognise that they were not well used and that microfilm targets would have served as well.

File format and file size – production aspects

Most file format issues were already decided outside the production process, but they did have some impacts:

Maintaining something like deliverable file sizes did constrain the vendors' options for solving image quality problems.

File naming became a production issue at various stages. The project decided to use "intelligent" file and directory names, built around ISSN numbers and incorporating dates, issue numbers, and page numbers. This was intended to make the files easier to track and easier to name; however, vendors found the naming convention complicated and prone to errors. An alternative "dumb" running number system may have been easier to apply. The National Library is currently exploring models for persistent names that can be applied to digital files required for ongoing public access. Its findings will influence naming conventions for future projects in which it is involved.

One advantage of the TIFF format is that metadata tags can be added to the digital object. The project specifications required the vendor to add certain information to the TIFF headers. This had some major production implications as vendors had to develop programs to add the information automatically. (There was also some discussion of the value of the information, given that most users, and indeed most project managers, did not have access to a viewer that would allow them to access or evaluate the header information.)

Evaluation of production

- Technical viability

The project often felt like it was on the edge of technical viability, trying to do things for which the technology was not yet available. And yet, the results suggest that most of what was wanted was achieved. The main technical issues remaining are:

- capturing sufficiently legible images from marginally legible originals – perhaps we need to explore alternatives such as direct scanning for some materials
- better tools for dealing with contrast problems
- techniques for dealing with broadsheet newspapers, probably including more effective compression approaches
- bandwidth improvements that will allow improved delivery speeds for large image files
- improvements in OCR technology to handle difficult historic print materials
- better tools for scanning from existing microfilm, to cope with the impact of skewed, inconsistently placed images. This remains an important goal, given the very large amount of microfilm that has been produced in the past and for which there would be value in more user-friendly access
- preservation techniques for digital files and technical infrastructure to manage them.

In almost all cases the technology already exists to address these issues.

- Preservation viability

The preservation viability of projects like ACDP is still to be proven. They provide at least a short-term preservation benefit as an alternative access medium to microfilm. However, the net effect is likely to be a considerably increased preservation requirement, to deal with the digital images as well as the print copy. However, the access benefits are such that similar projects will continue, their outputs becoming part of digital collections for which preservation programs are increasingly responsible.

At the other end of the process – ensuring material is not damaged in the digitisation process – the ACDP arrangements worked extremely well. The combination of a conscious and articulated commitment to protect the originals, well planned workflow arrangements, and carefully selected and conscientious vendors, produced ambitious results with no noticeable damage.

- Commercial viability

The commercial viability of the ACDP arrangements is harder to evaluate. The costs of digitisation are high, and the market in Australia remains small. From a purely commercial point of view it is unlikely that ACDP is repeatable. Its demand on resources from the project partners (and probably the vendors also) far exceeded expectations. Many processes that were expected to be easy to automate turned out still to require a large manual input. For the kind of difficult material processed in ACDP, a ‘production’ business model is probably quite different from the model that could be applied to less difficult material. We may need to devise hybrid models involving different balances of outsourcing and inhouse work where there are processes that are not commercially viable or over which we want to have a high degree of control.

On the other hand, it also must be recognised that ACDP was throughout its life experimental. There were large set-up costs and management inputs required at almost all stages. So long as future projects can make use of the many lessons learned from ACDP, and build directly on its foundations, the economic viability should improve considerably.

Conclusion

While in many ways the ACDP seems, particularly in hindsight, an ambitious - perhaps overly-ambitious - endeavour it has provided us with hard-learnt experience which is already being applied in other digital programmes. This project has assisted the National Library in assessing and furthering parts of its major Digital Services Project initiative and has provided the basis for further work. The State Library of NSW has initiated several digital projects, including the Banks and Flinders projects. The Library at Sydney University has established the Scholarly ElectronicText and Image Service (SETIS) which is developing several electronic text projects following the processing of the ACDP novels, and now hosts a large fulltext Australian literature database.

As importantly, the project has also provided us with conceptual understandings of digital library processes and structures as well as matters such as institutional infrastructure, organisational implications, new relationships with clients, the benefits and processes of collaboration, and - of course - concerns regarding costs and funding.

Digital conversion is not the seamless, simple process often portrayed by some digital library proponents. While the ACDP was funded to handle a set body of content, the trade-offs regarding cost-effectiveness and cost -benefits of digitisation requires some hardheaded and informed decision-making. The close management and scrutiny of the processes that has taken place throughout the ACDP has better equipped us for these tasks.

In the end, however, it is the content that validates the process, and the voices of the 1840s - ‘fully represented in the periodical press’ - are now transformed and freely accessible to give us an insight and understanding of their time and place.

ABOUT THE AUTHORS

Ross Coleman - Collection Management Librarian at the University of Sydney Library since 1991,. he has been involved in several collaborative digital library and electronic publishing projects.. He has been an initiator and manager for the ACDP 1840-45 since 1995, a manager of the AVCC-funded electronic Medical Journal of Australia (eMJA) projects, and through the Sydney University Library's Scholarly Electronic Text and Image Service (SETIS) has been involved in a number of projects including the Australian Text Database.

Colin Webb is Director of Preservation at the National Library of

Australia. He has worked as a bookbinder, conservator and preservation manager for the past 25 years. He vaguely remembers a time when his working life was filled with deteriorating library and archives collections of books, papers, and microfilm, and digital issues meant handling material with clean fingers. One of his current concerns is developing workable models for managing library preservation programs in a digital environment. He can be contacted at cwebb@nla.gov.au."

Appendix 1:
Tests and feedback

Tests and feedback phases in the project can be characterised as follows:

1st test (early 1996) – a small number of test files were produced by two vendors, using existing microfilm to create digital images in a range of resolutions and bit depths.

2nd test (mid 1996 – early 1997) – five titles, comprising c.7,600 pages of various dimensions, were digitised either from existing microfilm or from film created for the purpose. This test was subject to an Expressions of Interest process and detailed specifications. (For this phase the vendor used ScreenScan equipment to capture single images).

3rd test (late 1997) – on the basis of the 2nd test, specifications were revised and a Request for Tender was released for the production phase of the project. The first batch of material processed under this model was evaluated to test the effectiveness of the specifications. (For this phase the vendor use a Sunrise 35mm film scanner).

4th test (early-mid 1998) – because of serious image quality problems in Batch 1, a detailed assessment process recommended a further test phase. Initially, this involved scans of two titles using better scan settings, improved quality control by the vendors, and image enhancement software to clean the images.

5th test (mid 1998) – as these tests managed to produce cleaner but less legible images, a further test was conducted using adjustments in the software and in the scan levels.

6th test (late 1998) – on the basis of successful results in these tests, the vendors returned to production mode, although this was also treated as a test to see whether new scanning and post-scanning arrangements could deliver acceptable images at high output levels.

7th test (late 1998) – because the post-scanning file enhancement work was transferred from one vendor to another there was a further period of testing to ensure that the new arrangements would work.

8th test (early 1999) – production phase using new arrangements – necessitated further monitoring.

9th test (early 1999) – because file naming was partly transferred from one vendor to another there was a period of testing to ensure the new file naming arrangements would work.

10th test (early-mid 1999) – final production phase.

Appendix 2:
ACDP - facts and stats

- funded by the Australian Research Council Research Infrastructure Facilities and Equipment (RIEF) Program - partners: University of Sydney Libraries, National Library of Australia, State Library of NSW, Monash University Library

Journals at http://www.nla.gov.au/ferg/

- 61 titles included
- 58,213 pages scanned
- range of pages per title:
  smallest: 9 pages (South Australian Temperance Advocate)
  largest: 6,169 (Colonial Magazine and Commercial Maritime Journal)
- TIFF v.6.0 file format used for bitonal images at 400 dpi resolution relative to the paper originals
- total digital file size: approximately 14.33 GB
- range of total digital file size, (compressed), per title:
  smallest: 1.44 MB (South Australian Temperance Advocate)
  largest: 3.21 GB (Maitland Mercury)
- range of average file sizes (compressed) per title:
  smallest: 50 KB (Juvenile)
  largest: 4.9 MB (Geelong Advertiser)
- multi-page PDF format used for delivery over the Internet from National Library of Australia site (URL: http://www.nla.gov.au/ferg/)
- vendors involved: PSW Media, W&F Pascoe (microfilming, scanning, image enhancement, file naming); Discovery Media (image enhancement, file naming)
- equipment used by vendors:
  software used by vendors for image enhancement: inc. Kofax Toolkit, Equilibrium
  A variety of CD-Rs were provided by suppliers, primarily Kodak DS. Mitsui Gold were used for archival storage.
- CDs stored in environmentally controlled conditions, at 18 degrees C and 30% relative humidity
- images of serials were OCR'd using Adobe Fine Reader software
- 1021 image files were made of the Geelong Advertiser 1840-41, captured directly by digital camera by Australian Conservation Services

Novels at http://setis.library.usyd.edu.au/acdp/

- 4 fiction titles included
- vendors W&F Pascoe (microfilm and scan), OCR and Proofreading: Anamaria Beligan The Athanor P/L (Melbourne)
- Creation of TEI.2-conformant markup: Creagh Cole Scholarly Electronic Text and Image Service (SETIS), University of Sydney Library
- these titles formed the genesis of the Australian Literary and Historical Texts collection at http://setis.library.usyd.edu.au/ozlit/