# The 'Banks on CD-ROM' project at the State Library of New South Wales

*Louise Anemaat*
*with Paul Brunton and Alan Ventress*

**Louise Anemaat** is Project Archivist on the 'Banks on CD-ROM' project. She has worked in the Manuscripts Section of the Mitchell Library since completing the Graduate Diploma in Archives Administration in 1988. As the 1994 recipient of the Esso Scholarship for the professional staff of Australian libraries, galleries and museums, she spent part of 1994 based in Berlin, locating and researching original material, and researching German archives practice.

**Paul Brunton** is the Curator of Manuscripts, Mitchell Library, and immediate past President of the Australian Society of Archivists. **Alan Ventress** is Manager, Australian Research Collections and Mitchell Librarian, and is currently working on projects to digitise some of the State Library's Australiana collections. Both were members of the original project team, which also included Margy Burn, Director, Information and Collection Services and John Murphy, Senior Manuscripts Librarian.

*A project to digitise the papers of Sir Joseph Banks held in the collections of the Mitchell and Dixson Libraries, and make them available as indexed images on CD-ROM, is currently underway at the State Library of New South Wales. The extensive manuscript collection contains correspondence, sketches, plans and journals of Banks dating from 1768 when he left England on the* Endeavour *voyage which took him to the east coast of Australia, until his death in 1820. The CD-ROM will contain around 10 000 facsimile images of the papers, indexed such that the user can gain quick access to a single document or to a series of documents.*

PHILIP LARKIN, IN A LECTURE DELIVERED IN 1979, referred to literary manuscripts as having two kinds of value—the magical and the meaningful.[1] These values could equally be applied to historical manuscripts such as the papers of Sir Joseph Banks held in the collections of the Mitchell and Dixson Libraries at the State Library of New South Wales. For Larkin, the magical value is the older and more universal, it is the paper that has been written on, the words as they were written, changes that were made, the particular combination of words and ideas; the meaningful value includes the degree to which the manuscript helps to enlarge our knowledge and understanding of the writer's life and work.

One of the challenges of the State Library's current digitisation project, the 'Banks on CD-ROM' project, has been to preserve and faithfully reproduce both the magical and meaningful values of the Banks papers for researchers. This challenge has coloured every decision in the long process, currently underway, of preparing a large and significant personal archive for publication on CD-ROM.

The project is basically about improving intellectual access. The greater searching capacity of the technology allows the papers to be interrogated in a way not possible previously. The project has also been an opportunity to experiment with the application of a new technology to archives and manuscripts collections. Most imaging projects to date have concentrated on printed material, or single items or treasures, rather than on interrelated accumulations of documents.

## The Banks papers

That portion of the papers of Sir Joseph Banks (1743–1820) held in the State Library of New South Wales contains some of the most significant Australian colonial material held anywhere in the world. Despite some obvious gaps, enough is held in the State Library's collection to gain a definite sense of the scope and variety of Sir Joseph Banks' interests and activities and the extent of his commitment to them.

From the time of James Cook's *Endeavour* voyage, on which Banks travelled as a botanist, until his death in 1820, Banks was actively involved in almost every aspect of Pacific exploration and early Australian colonial life. His papers document not only the *Endeavour* voyage which Banks joined (1768–1771), but also Cook's later voyages (1772–1775 and 1776–1780). Banks actively supported the proposal of Botany Bay as a site for British settlement. He proposed William Bligh to command two voyages (1787–1790 and 1791–1793)

including the ill-fated voyage on the *Bounty* which ended in mutiny. He later recommended Bligh to succeed Philip Gidley King as the fourth Governor of New South Wales. Bligh's governorship ended in his deposition in the Rum Rebellion in 1808. Banks corresponded with the first four governors of New South Wales who, while they reported officially to the Secretary of State for the Colonies, also reported privately and therefore more intimately and openly to Banks.

Banks organised Matthew Flinders' voyage in the *Investigator* (1801–1803) which defined the map of Australia. He had connections with Sir George Macartney's embassy to China (1792–1797) and with George Vancouver's voyage to the north-west coast of America (1791–1794). He sent botanists to all parts of the world, including New South Wales, often at his own expense.

Practically anyone who wanted to travel to New South Wales, in almost any capacity, consulted Sir Joseph Banks. He was the one constant throughout the first thirty years of white settlement in Australia, through changes of ministers, government and policy. The importance of his papers lies in the light they throw on the reasons for white settlement in New South Wales, on the environmental impact of settlement, the impact on and attitudes to indigenous peoples, civil and military power relations and the development of constitutional government.

## Funding

The project has been made possible by the Sir Joseph Banks Memorial Fund which dates back to 1905 and was the initiative of J. H. Maiden, FRS. Funds were raised through public subscription and the sale of Maiden's book, *Sir Joseph Banks: the father of Australia* (1909) and have been vested in the State Library since 1946. They were to be used 'in or towards defraying the cost of editing, publishing and distributing the Banks papers in a manner and form suitable and fitting to the memory and services of Sir Joseph Banks'.

The Library previously used the Fund to publish Banks' journal from the *Endeavour* in 1962, to subsidise Averil Lysaght's *Joseph Banks in Newfoundland and Labrador* in 1971, and to support the publication in 1979 of Banks' correspondence, mainly held in the Sutro Library, California, relating to sheep and wool.

## Acquisition of the papers

The papers of Sir Joseph Banks at the State Library are held in the collections of both the Mitchell and Dixson Libraries, and have been acquired in many consignments over a period of more than 100 years.

The Mitchell Library, opened in 1910, is based on David Scott Mitchell's personal collection of 61 000 books, and thousands of manuscripts, paintings, photographs and maps relating to Australia and the Pacific bequeathed to the nation in 1907. Today the collection comprises around 9 000 metres of manuscripts, over one million photographs and pictures, hundreds of thousands of maps and more than half a million printed items in diverse formats. The smaller personal collection of Sir William Dixson, covering the same collecting areas as Mitchell's and bequeathed to the state in 1952, constitutes the Dixson Library which opened to the public in 1959.

When Banks died, he left behind a well organised archive which documented his influential career. It was the most comprehensive archive of its kind in Britain—perhaps the world.

The custodial history of the archive is confused. Most of it was eventually deposited in the British Museum, probably around 1873, only to be removed in 1880 by Edward Knatchbull-Hugessen, Lord Brabourne, a collateral descendant of Banks. Brabourne successfully claimed that the papers had only been lent to the British Museum. The papers were returned to Brabourne who subsequently offered the papers for purchase to the British Museum for £250. The Museum declined. Then began the dismemberment of the archive which resulted in the loss of much information about Banks, his life and period. Brabourne started selling the papers. His descendants continued.

J. C. Beaglehole, biographer of James Cook and editor of Banks' *Endeavour* journal wrote: 'Thus was one of the greatest collections of records for the history of science and the social life of the intellect ever assembled, a collection fundamental for our knowledge of certain aspects of the eighteenth century, flung away and dissipated all over the earth'.[2]

In 1884 the Agent-General for New South Wales in London, Sir Saul Samuel, acquired from Brabourne most of the Banks papers relating to Australia and the Pacific for the sum of £375. Purchased for the state, this was a personal act of far-sighted patriotism on behalf of Samuel unprompted by the New South Wales Government. The payment was also to cover any further Australian

papers of Banks which might come to light though this commitment, which Samuel had in writing, was not honoured.

The papers did not come to the Library immediately. Initially they were used in the compilation of the *History of New South Wales from the records*, the first volume of which appeared in 1889, published by the New South Wales Government Printer. They were again used in the multi-volume *Historical records of New South Wales*, which appeared from 1892 until 1901, also emanating from the Government Printer. The papers, having already been subject to earlier reorganisation, were annotated and rearranged during this process. They were transferred to the Mitchell Library when it opened as the Australiana collection in 1910 and have become known as the Brabourne collection, an appellation which obscured Banks.

Banks papers, including the *Endeavour* journal, were also included in Mitchell's bequest. Mitchell had purchased these from other collectors who had acquired them, perhaps at several removes, from Brabourne.

The next major acquisition was in 1929 through a Sotheby's auction in London. Banks' stocks had risen and the Library had to pay £7 000 for approximately half the quantity of papers purchased in 1884. Most of these papers were in volumes assembled at least by the Brabourne family, if not by Banks himself. However, for reasons which are now unknown, it was decided that some of the letters should be extracted from the volumes and these were acquired by the National Library of Australia, archival integrity not being a governing principle.

In the ensuing years, numerous smaller accessions have been made as estrays from the now far-flung Banks archive appeared in auction sales and dealers' catalogues. The latest acquisitions have been a single letter written by Comte de Lauraguais, acquired in 1989 for $7 500, and most recently, a single letter written by George Caley, acquired this year for $850.

## Organisation of the papers

At the inception of the project, the collections of the Mitchell and Dixson Libraries were comprehensively surveyed for all material of Banks provenance including manuscripts, pictures and maps. The final collection, separated from misattributed material or items of other provenance, amounted to approximately 10 000 pages and included correspondence, principally letters received, but also reports, accounts, journals, plus a small quantity of maps, charts and watercolours. The Banks project will include only those papers of

Banks' provenance which are held by the State Library of New South Wales though the possibility now exists of reuniting the entire archive electronically, despite being spread around dozens of institutions worldwide.

Given the existing arrangement and condition of the papers it was considered vital to unbind them and to also take the opportunity of performing some restoration. This work, which has been carried out in tandem with arrangement and description work, has naturally proved very time consuming and is only now nearing completion.

The papers are being rearranged to reflect the way they were used and accumulated by Banks. In some cases this has been fairly straightforward, in others very complex. We know Banks kept together papers relating to specific activities, often with assigned titles, a practice which he began quite early. An early file title included in his papers, but unfortunately with the contents missing, reads 'Letters containing Science written Prior to the month of August 1768', which coincides with the departure of the *Endeavour*. A later note written in 1799 in Banks' hand reads:

> if these Collections which now consist of 15 volumes shall by any accident descend to Posterity the Collector requests that allowance may be made for numerous Letters which will be uninteresting to the Reader & Recollect that numerous tempting reasons occurred which Compelld him to preserve for his own sake matters not likely to interest anyone else he hopes however he has done his duty to his successors by destroying nothing likely either to interest or amuse his successors.

Similarly, papers relating to the equipping of the *Bounty* voyage of William Bligh are grouped together under Banks' own title 'Plan for the Voyage with Letters from various persons who interferd in the management of it'. Letters Banks received from Bligh during the voyage of the *Bounty* have been grouped together, by Banks, under the title 'Correspondence Bounty'. Papers relating to the equipping of the *Investigator* voyage under Matthew Flinders are grouped together under the heading 'Correspondence relating to the fitting out of the Investigator for a voyage of discovery'. He kept these separate from another series titled 'Correspondence with Matthew Flinders'.

These archival series will of course remain in their current arrangement. The only changes will be the addition of some fugitives which had been removed by previous owners and can now be reinserted.

Banks evolved into a diligent, disciplined correspondent in response to his manifold interests and activities. He received letters from diverse people, from

all over the world, covering a vast range of subjects. He wrote up to fifty letters a week. Of his replies, there are relatively few examples in his collection. If they are included, his replies are usually in draft form or in summary at the end of the letter received. Occasionally a complete letter is included, usually in the hand of his clerk, William Cartlich.

In response to such a bulk of correspondence, it seems probable, from information contained in several documents in the collection titled 'Index to names of writers', that Banks often numbered his correspondence and arranged it in volumes alphabetically by correspondent's name, and chronologically in order of receipt, within that. A volume covered a given year. Many of the letters he received from his extensive correspondents also include a folio number from these volumes, written in ink in the top right hand corner. Sometimes a single document will contain three or four folio numbers variously imposed by Banks or a clerk, the Government Printers, the Library or previous owners.

It is difficult to determine how far this arrangement extended across Banks' entire archive. For this reason it is also not possible to replicate entirely Banks' original arrangement. A decision was taken therefore to arrange letters not clearly part of a series devised by Banks into series based on correspondent and arranged chronologically within that. This practical arrangement approximates Banks' probable original arrangement.

Parts of the papers have become known to researchers by their shelf location numbers which have appeared as published citations and which are no longer relevant in the new arrangement. The previous shelf location number of every document has been recorded for inclusion on the CD-ROM.

## Description

On the surface at least, it seemed there were various options for the type of CD-ROM product which would result from the project including the transcription of the papers. This had some appeal because of the potential for full text retrieval of the documents. The cost, however, proved prohibitive—around $50 000 for an unedited first draft. Transcription would also marginally increase the storage requirements because both the original document in facsimile and the transcription would need to be scanned. Transcription is the option taken by a separate digitisation project with which the State Library has had useful contact. This project, at the University of Bergen in Norway involves digitising the archive of the philosopher Ludwig Wittgenstein.

Optical Character Recognition (OCR) was investigated but was never a real option for a manuscripts digitisation project of this nature as it is inadequate for capturing both the information (Larkin's meaningful value) and the presentation of the original manuscript page (the magical value) which is critical. With current standards, OCR simply could not be considered for the digitisation of eighteenth and nineteenth century manuscript items encompassing many different styles and standards of handwriting and eighteenth century spellings. This may become an option in the future as machine recognition of handwriting improves and becomes more commonplace.

A multimedia style product was considered but rejected, despite the perceived potential commercial appeal, on the grounds of high cost (between $100 000 and $150 000), lack of appropriate expertise within the Library and, importantly, the concern that the archival integrity of the collection would be compromised by selecting edited highlights of the papers rather than reproducing the entire archive.

This left the only realistic option of manually creating a bibliographic database for the entire collection based on two record types—one describing and indexing the series; the other describing and indexing each individual document within a series.

A PC based retrieval system was eventually adopted. It was also decided that the final product should run on a Windows based PC using a Graphical User Interface (GUI). The possibility of running the system on different platforms using a 'hybrid' Mac/Windows product was also investigated but abandoned due to the limited storage on a single CD-ROM and the high cost of dual interfaces.

The **series level description** includes:

- **Series title**, with date range;
- **Provenance note**, which describes the custodial history of each document in the series, including previous Mitchell and Dixson Library shelf location numbers;
- **Background note**, equivalent to a Biographical Note or Administrative History; and

- **Subjects**, which involves indexing the series for persons' names, ships' names, events, and other topics, common to the entire series.

Very few documents can be fully understood by themselves; it is their accumulation into a meaningful context—the series—which provides the subject. The purpose of these entries, based mainly on Australian Bibliographic Network authorities, used by the *Guide to collections of manuscripts relating to Australia*, is to describe the common content of each document within the context of the series.

In addition to the series level description, **document level description** of each individual document in the series is included. This is the difference between our usual level of description and indexing, and the level required for this project. Each document description includes:

- **Document title**, including the date of the document;
- **Author** of document;
- **Date** of document;
- **Series title** to which the particular document belongs;
- **Language**, if other than English, or English and a second language; and
- **Subjects**, including additional persons' names, ships' names, events, and other topics, which are unique to the document described and which therefore do not duplicate subject entries which occur at the series level.

All the fields, at series and document description level, are free-text searchable. Other features include limited Boolean searching, and the ability to annotate to a word processor. It is also anticipated that users, typically engaged in slow, painstaking study and analysis, will want to make hard copies of documents from the CD-ROM.

This very detailed level of description and indexing is the minimum level considered necessary to take advantage of the high speed retrieval and multiple simultaneous access capabilities of CD-ROM technology. If anything less than this level of access were provided, the benefits of reproducing a collection on CD-ROM would be barely greater than those gained by producing a high quality microfilm. An institution considering this type of project as a means of providing access to collections must be properly informed and aware of the high level of commitment, particularly in terms of staff time,

which is required, and must decide whether this can be justified by the outcome.

## Software interface

In terms of the interface, a number of options was explored including off-site development by an outside agency, a combination of off-site and in-house development, and total in-house development. It was decided the most useful approach was a combination of off-site and on-site development to both take advantage of professional expertise not available in the Library and to build up staff knowledge of digitisation in-house.

Products investigated include a proprietary solution produced by Info One; the possibility of utilising Sanderson's 'First' Software; Inmagic Plus Image Version for Document Management; Hyper-Writer from CD-Networks; ISYS from Discovery Media, the producers of the Australian Visual Arts Database on CD-ROM; and Microsoft Multimedia Viewer, also in association with Discovery Media.

The importance of selecting the best retrieval engine available cannot be over emphasised. After assessing all products in terms of user friendliness, cost, royalties and availability in Australia, Microsoft Multimedia Viewer was chosen because it was royalty free, was widely used (MS-Encarta, MS-National Gallery, MS-Cinemania), and had been donated to the Library by Microsoft Australia.

## Imaging

The Banks project took the option of scanning the manuscripts from microfilm rather than exposing the documents themselves to the scanning process. This decision was made for several reasons including the need to provide readers who may request it with a microfilm copy, and to more easily accommodate, during the scanning process, paper sizes which vary within the collection from very small scraps of paper to A3 pages or larger.

Both the microfilming, done to the level of 160 line pairs per millimetre, and the digitisation are being carried out by the Library's usual microfilm contractors, W. & F. Pascoe. Pascoes generate the original G4 TIF files, uncompressed, as well as the compressed and adjusted JPEG files which are then supplied to Discovery Media. Each image file averages 400 kilobytes after scanning in and compression.

The decision to microfilm was also taken because of the reality of equipment obsolescence and the consequent need for a human readable copy, and difficulties with standards. While CD-ROMs are fairly well established and standardised, problems with the disc as a physical product still exist, as Frederick Stielow[3] has outlined. Variations occur from manufacturer to manufacturer and within a batch from any given manufacturer. They may be subject to physical and chemical changes due to a variety of environmental conditions, problems in manufacturing or damage in handling. There have also been problems reported with fungus, disc rot, label ink eat-through etc.

With respect to longevity, Stielow points out that claims made by the industry have escalated from early ten year predictions to Sony's or Digipress' more reliable Century Disk with its 100 year predicted life. Stielow's more conservative estimate seems to be within the range of 10–25 years.

The safest assumption when deciding on CD-ROM or other imaging formats for heritage materials is that the available optical storage media have limited archival properties and cannot be relied on exclusively for the permanent preservation of important materials. On the contrary, information in published studies and manufacturer's product literature suggests that optical storage media are relatively unstable when compared with paper and microfilm.

Advantages accrue in respect to copying digital media. Unlike the copying of magnetic storage media, such as cassette and videotape, or the copying of microfilm, in which each new generation or recopying produces a diminished copy, digitised records can be copied without generational loss.

Because it is such a rapidly evolving and developing technology, however, the media longevity questions raised by Stielow may not even present themselves as problems. Long before the media deteriorate, the equipment it runs on will have been rendered obsolete by new systems. The scanned and stored images for the Banks project and the bibliographic database may, of course, be transferable for application to future media.

Another important variable when considering digital image preservation is the resolution of the image measured as dots per inch (dpi). The greater the dpi the clearer the image but the greater the storage requirements, unlike microfilm which is resolution indifferent. As image storage is one of the major capital costs of digitisation, the images will be stored compressed.

For the Banks project it was decided to scan at a minimum resolution of 100 dpi and using 256 grey scales rather than simply black and white. The

small amount of pictorial material in the Banks collection, mainly watercolours, will be scanned from full colour microfilm. Where a single document in colour occurs in a series, the entire series will be microfilmed in black and white as if the colour document did not exist. The colour document will then be separately scanned in full colour. The implication of this, however, is that some images included on the reference microfilm will differ slightly from the digitised product. An alternative would have been to microfilm the coloured document in colour and splice this image into the black and white microfilm before scanning the images.

Various degrees of limited enhancement of the image will occur during the scanning process to reduce verso bleedthrough, staining or other residues etc. Over enhancement of heritage materials raises the issue of the authenticity of the document as a trade-off to document legibility—magical versus meaningful value. A digitisation project of Civil War documents in USA did rely heavily on digital enhancement to improve the legibility of documents which had been written in ink diluted with water, a response to chronic wartime shortages. While this may be a valid option, it unquestionably changes the effect the documents create and destroys some of the information implied and conveyed.

Limited possibility also exists for some image adjustment at the user interface. In addition to standard functions for displaying and scrolling through documents, and zooming in to documents, users may also employ tools to increase or decrease contrast and brightness.

## Client council

An early decision was made to create a 'proof of concept' disc using approximately 300 images and the associated index to enable staff and potential users of the Banks papers on CD-ROM to make a well informed decision about the final product. The prototype was presented to a client council, which included historians, a teacher, genealogist and a local history librarian, in May of this year.

The prototype was also sent for comment to Australian Archives, the National Library of Australia and the Banks Archive Project at the National History Museum, London.

## Conclusion

Some of the value of the Banks project is undoubtedly in the opportunity it provides to subject the Banks papers to a thorough analysis in an attempt to replicate the original archive as far as possible. The very close and detailed description required and the recreation of series has led to the identification and correction of long perpetuated misattributions of documents and incorrect datings.

Each collection is different bringing with it a new set of problems and challenges, especially in arrangement and description. The arrangement of the Banks papers has been very complex because of the way the papers were acquired and used in the more than 170 years since Banks' death. Description has been one of the greatest challenges and has placed the greatest demands on the project.

While the Banks papers are unusual in their complexity, disorder and the partial double provenance introduced by their use by the Government Printers, and while each new collection differs from the last, the approach taken by this project suggests a workable basis for other similar projects based on archival materials. However, digitisation of a collection which had come to the archivist in its original order would have been a more straightforward proposition for a first imaging project.

Following his voyage with Cook in the *Endeavour*, Banks embarked on one of the most ambitious publishing ventures of all time. The botanical results of this voyage were such that they revolutionised European knowledge. Banks had intended to disseminate this knowledge and had over 700 copperplates of new plants engraved and their descriptions prepared.

Owing either to lack of money, or to the death of Daniel Solander, his long-time friend and librarian, or to his involvement in other projects, this publication never appeared in Banks' lifetime and the copperplates remained unused. The information was disseminated by others who visited Banks and used the records. Eventually, in the 1980s, these copperplates were used to publish the entire botanical history of the voyage as Banks had intended.

Our work on the Banks papers in order to publish them on CD-ROM is similar to the work in producing the copperplates. We have produced scanned images of the documents and married these with extensive indexes—our copperplates, if you like. At present the intention is to disseminate this information by means of CD-ROM. However, when this technology is

superseded it is hoped our copperplates can be adapted and used by the preferred carrier of the day.

While there is currently a lot of interest in and excitement about the possibilities of digitisation and CD-ROM technology, it by no means presents archivists with a panacea. It is a challenging interim in the technological continuum which, like any new technology, compels archivists to re-examine their objectives, theory and priorities, and to meet new preservation demands.

Digitisation is still expensive and time consuming, and reading from a computer screen is still not as comfortable as reading from a printed page but, if used selectively, it is a process with the potential to enhance intellectual access to our collections. We look forward to the anticipated completion of the Banks project early next year, our, and we believe Australia's, first venture into the digitisation of an archival collection.

## Endnotes

1. Philip Larkin, quoted in Michael Holroyd, 'How do we block the drain? The problems of locating and retaining literary papers', *Times Literary Supplement*, 23 June 1995, p. 24.
2. J. C. Beaglehole (ed.), *The Endeavour journal of Joseph Banks, 1768–1771*, vol. 1, The Trustees of the Public Library of New South Wales in association with Angus & Robertson, Sydney, 1962, p. 137.
3. Frederick J. Stielow, 'Archival Theory and the Preservation of Electronic Media: Opportunities and Standards Below the Cutting Edge', *American Archivist*, vol. 55, Spring 1992, p. 332–343.