

STAFF PAPER



## Transforming access to Australia's documentary history

The 2007 Ferguson Memorial Lecture  
At the invitation of the Royal Australian Historical Society  
Ferguson Hall, St. Stephen's Uniting Church, Sydney  
29 August 2007

Warwick Cathro  
Assistant Director-General,  
Innovation  
National Library of Australia  
wcathro@nla.gov.au

## **Introduction: Sir John Ferguson**

Members of the Ferguson family; members of the Royal Australian Historical Society; ladies and gentlemen: it is a privilege for me to give this lecture in honour of Sir John Ferguson, who stands in the first rank of Australia's great collectors and bibliographers.

We are gathered in a hall named in honour of Sir John's father, the Reverend John Ferguson, who was Minister here at St. Stephen's from 1894 until his death in 1925. It is a good place to think about the life and achievements of Sir John Ferguson, who was inspired by his father with a love of books and scholarship.

Late in his life (in the year 1962) Sir John Ferguson made an oral history recording with Hazel de Berg for the National Library. In that talk he recounted his origins in Invercargill, New Zealand, his move to Sydney at the age of 12 when his father was appointed to St. Stephen's, his education at the University of Sydney, his career at the bar, his subsequent career as a judge of the Industrial Commission of New South Wales, and of course his two terms as President of the Royal Australian Historical Society.

Sir John was also the son-in-law of George Robertson the publisher and bookseller, a fact which enhanced the interest in books that he had acquired from his own father. He tells Hazel de Berg how he spent many hours wandering through the second hand bookshops of Sydney and other places. As he put it, his pursuit of books was undertaken "in small shops, in unsuspected resorts, up narrow lanes, in old towns, in furniture stores [and] in auction rooms".

In this lecture I will aim to describe how the National Library of Australia is honouring its debt to Sir John Ferguson and how we are responding to changes in society and technology as we interpret our obligation to enhance access to our documentary history.

## Ferguson's legacy and the National Library of Australia

Located on the mezzanine floor of the National Library, looking down on the foyer, is the Ferguson Room. It honours the memory of Sir John Ferguson, and contains a display of material related to the Ferguson Collection. It is a living space which is used almost every day for the education of both readers and library staff.

The room is a demonstration that the National Library will never forget the legacy which it has inherited from Ferguson and indeed from other great collectors like Petherick and Nan Kivell. Ferguson himself, speaking of collectors like Mitchell and Petherick, commented: "Those great collectors rendered an important service by making us Australiana conscious and by saving for our nation many treasures which would have inevitably perished".

This year marks 70 years since Sir John offered his collection to the National Library. Of course it was not officially called the "National Library" in 1937, but the then Librarian, Kenneth Binns, had been using that name in a semi-official way for several years. Why did Ferguson make that choice, given that he was also a trustee of the Public Library of New South Wales? In "Remarkable occurrences", a book commemorating the National Library's first 100 years, our former Manuscript Librarian Graeme Powell explained it as follows:

*Although he spent all of his life in one city, Ferguson was a nationalist and in his writings frequently referred to the development of the Commonwealth. He admired the dedication of Binns in building up the collections of the Commonwealth National Library.... He argued that his collection would be of more value in Canberra than in Sydney, as a large proportion of his published holdings would already be in Mitchell's collection.*

This is an appropriate point to pay a tribute to our colleagues from that great institution, the State Library of New South Wales. It is perfectly natural for a spirit of friendly rivalry to exist between great libraries in the collecting sphere, but the emphasis these days is very much on the "friendly" rather than the "rivalry", and as someone who has worked at the National Library for almost 30 years, I can assert that the level of engagement and cooperation between our two libraries has never been stronger than it is today.

At the National Library we are still following in the Ferguson tradition. We continue to seek out retrospective Australian printed material in order to complete a comprehensive collection of nineteenth and twentieth century imprints. One of our top priorities is to collect material listed in Ferguson's *Bibliography of Australia*, or not listed but eligible for inclusion in it.

We are still having success in this endeavour. To give a couple of examples:

- Last year we acquired John Lhotsky's *A journey from Sydney to the Australian Alps*, a book published in the 1830s for which only one other complete copy is known, and that copy is held by the British Library.
- Last month we acquired another Ferguson item, W. J. O'Donnell's *Diary of the exploring expedition*, a rare pamphlet published in 1884, describing an exploration of Australia's north-west coast.

These publications were acquired at the first and third auction sales of the collection of the Melbourne based Australiana collector Rodney Davidson. At the third auction, which covered the publication period 1851-1900, there were 213 lots. The National Library already held 200 of them, so we bid for just 13, and we were successful in acquiring 11 of those. So you can see that we are far advanced in our goal of a comprehensive collection of nineteenth century Australiana.

Even with more recent publications, we are on the lookout for material which may have escaped the legal deposit net, such as material from private presses and small publishers, limited editions and significant ephemeral items.

## Digital collecting

In recent years the National Library has responded to a clear need to interpret our collecting mandate in the light of great changes that have been occurring in the way that knowledge is recorded and communicated.

Sir John Ferguson said, in an address to the Royal Australian Historical Society in 1922:

*It should be the aim of history to bring the past into manifest relation to the present and to show how historical ideas and experiences are being constantly projected into the present.*

Any serious memory institution like the National Library of Australia must consider how to give life to this goal by imagining the experience of a future user, somebody investigating what Australia was like in the distant past of 2007.

For the past 15 years or so, most writing has been created in digital form. And now an increasing proportion of writing is distributed only in digital form, not existing in print form unless a user happens to print it locally after downloading it from a web site. Our libraries must respond to the challenge of collecting these so-

called “born digital” publications if we are serious about transmitting knowledge of the present to the future. The most important response involves making archive copies of web sites.

This is what the National Library and our partner institutions (mainly the state libraries) have been doing for the past 11 years: making, for posterity, archive copies of web sites. The copies can be found in PANDORA, Australia’s web archive, which is hosted by the National Library.

The PANDORA archive is heavily used. In fact, our digital collections as a whole are heavily used, with about 13 million online accesses each year. Contrast that with the 300,000 uses per year of our physical collections.

In building the PANDORA archive we are driven by two major imperatives. First, we know that many significant web sites will disappear relatively quickly from the Internet when the funding that supported them ceases, or when the activity that they reflect ends. Thus, for example:

- The official web site for the Sydney Olympic Games disappeared from the Internet in 2001. However, the National Library archived the complete web site 26 times – including every night during the Games.
- The Centenary of Federation web site, produced by the National Council for the Centenary of Federation, disappeared from the Internet at the end of 2002, along with its rich set of information resources about the history of Australia. However, the National Library archived it four times. The archive snapshot taken on 3 January 2001 contains goodwill messages from the heads of state of more than 80 countries on the occasion of our centenary as a nation. The Centenary of Federation snapshots continue to be among the most heavily used of all of the sites archived in PANDORA.
- The Interfet web site, recording the activities of the international peace-keeping force in East Timor in 1999-2000, has disappeared from the Internet. But the National Library archived it five times, to preserve a record of that historic peace-keeping enterprise.
- Most web sites created for Australia’s state and federal election campaigns since 1996 have disappeared from the Internet. But the National Library and the state libraries have captured them and they are still available through PANDORA.

Even where web sites still exist, they usually change significantly over time in their appearance and their content. A good example is the web site of the Australian Republican Movement. This site has been archived 16 times since July

1998, including an historic capture taken just before the referendum of November 1999. The archive snapshots show major changes in the content, design and even the mood of that site.

Our first imperative was to capture important web sites before they disappear or change. Our second imperative is to deal with the fact that all digital information is at risk of becoming inaccessible due to technical obsolescence. Let me give an illustration.

The National Library recently received a collection of papers and research data from Dr Robin Warren. Dr Warren is the Australian pathologist who in 2005 was awarded the Nobel Prize for Medicine, jointly with his colleague Dr Barry Marshall, for their work on the causes of stomach ulcers. Some of Warren and Marshall's key research was undertaken in the late 1970s and early 1980s, not long after the invention of the personal computer. Their data was stored on 5¼ inch disks which started to become obsolete in the late 1980s. The recovery and preservation of this data will be a significant challenge. Even though 5¼ inch disk drives can still be found, the file format standards of the 1980s are no longer in use, so interpreting and processing the raw data will be difficult.

Digital resources are usually created without long term permanence in mind. During the brief period of 30 years since the personal computer was invented, the methods we use to access and navigate information on those computers has changed dramatically. What will the computer interfaces of 2107 be like, and who is to say that these interfaces will work with the digital information created today?

We will respond to this challenge by migrating data to new formats before obsolescence occurs. We will routinely monitor our archive to determine whether any file formats are at risk of obsolescence. Librarians and digital archivists are now building tools to alert us to format risk and obsolescence. At the National Library we are already working with others to help to build and improve those tools.

The PANDORA archive is comprised of a selection of Australian web sites – a selection which is made according to well defined and published criteria, but one which nevertheless represents only a tiny proportion of the whole Australian web domain. There is a high probability that web sites which would be valuable for research in the light of hindsight will not be archived through the PANDORA approach.

For this reason the National Library has embarked on another approach to web archiving which complements the PANDORA approach. We call this the “Whole Domain Harvest”. In 2005 and 2006 we commissioned an institution based in California (the Internet Archive) to collect every Australian web site that it could identify, and to send us the archive copies. In the 2006 harvest we received over 500 million files comprising 19 terabytes of information. That is like 250 modern personal computers with their disk drives full to capacity. Or, if you like, the equivalent of more than 10 million of the old 3.5 inch floppy disks. It’s quite a lot of data. And remember that we intend to do this once per year, that every time we do it the size of the snapshot will be bigger, and that we intend to retain all of the historic snapshots.

We will preserve these snapshots, but in the short term we will not be able to open them up for public access, because (in contrast to PANDORA) we do not have explicit permission to do so from the millions of web publishers involved.

For the Whole Domain Harvest to continue in a satisfactory way, the Library needs two things to happen. First, we need changes to the legal deposit provisions of the Copyright Act so that our right to make these copies is placed beyond doubt. That Act currently provides that publishers must deposit one copy of any “library material” published in Australia with the National Library, but the definition of “library material” limits the requirement to items in printed form.

Second, the Library needs a statutory right to provide reasonable public access to our archive copies where this does not jeopardize the economic interests of publishers. In practice we would like to be able to provide such access immediately where the live publication is freely available on the Web, and in other cases provide such access after an agreed period of commercial availability has elapsed.

### **Oral history – collecting and delivery**

I have quoted already from the 1962 recording made by Sir John Ferguson that is held in the National Library’s oral history collection. The Library has a collection of more than 16,000 oral history interviews comprising more than 40,000 hours of recorded sound. The collection includes interviews with politicians, writers, sporting personalities, prominent professionals and ordinary Australians from many walks of life. Our aim at the Library is to preserve this collection, to continue to expand it, and to provide better access to it.

More than half of this collection is still held in analogue (ie non-digital) tape format. This older part of the collection is at great risk of becoming inaccessible due to the disappearance of the equipment needed to play analogue sound recordings. Consequently the Library has embarked on a 15 year program to transfer the analogue sound recordings to digital format, and to store them in our digital storage system. We are aiming to complete this task by 2018. The recording by Sir John Ferguson was converted to digital form earlier this month.

We also wish to make it much easier for researchers and the public to gain access to our oral history collection. Our aim is eventually to deliver on the Web every interview in our collection that is not subject to access restrictions. A key to doing this will be to use the transcript or summary of the interview, where such a text form of the interview exists. In the case of the recording by Sir John Ferguson, the transcript comprises 17 pages that were typed in the era before computers. We need to convert it to computer-based text and structure it so that we can link sections of the text to time markers in the digital audio file.

When we have done that (along with a few other technical procedures) it will be possible to deliver this interview on the Web. That means that anybody searching the National Library's database for anything about (say) David Scott Mitchell would turn up, among other things, the interview with Ferguson. The search would take you to the part of the transcript where Ferguson says that "Mitchell consciously set out to gather anything and everything about Australia .... He once justified the acquisition of a worthless publication by the reply that its presence in his collection was necessary in order to prove its worthlessness". One more click, and you will be able to play the relevant part of the audio file and hear Ferguson utter those words.

## **Bibliography**

Sir John Ferguson was renowned for his achievements, not only as a collector, but also as a bibliographer. He started work on his *Bibliography of Australia* in the 1920s and in 1941 published the first of what was to become seven volumes.

Writing in "Remarkable occurrences", Graeme Powell comments as follows:

*Ferguson became almost a compulsive bibliographer .... His bibliographies represented a considerable advance on Petherick's bibliographies. Their scope was well defined and, as well as providing full imprints, they described the physical features of the works in detail, including line breaks in the titles, paginations, illustrations and precise size.*

Bibliography is all about revealing and describing what has been published, and arranging that information so others may access it.

At the National Library we are engaged in a never ending labour of bibliography that has built on the work of Sir John Ferguson and will extend it indefinitely into the future. We aim to ensure firstly that all 6 million items in our own collection are catalogued at some level, and secondly that these catalogue records are accessible online to support the information needs of scholars and the general public.

We have catalogued more than 89% of those 6 million items, and almost all of those catalogue records are accessible online. The remaining 11% are mostly pictures and maps. Each year we continue to push that online figure up towards our goal of 100%.

Our experience is that the use of segments of our collection (say, French revolutionary pamphlets, to take an actual example) escalates dramatically after we have invested effort in the cataloguing and online exposure of those materials.

This bibliographic information does not get created by magic. Sir John Ferguson laboured for decades over his bibliography, and Australian libraries have collectively invested many millions of dollars in the task of cataloguing their collections.

Some of these projects can take many years to complete. As an example, the National Library holds over 175,000 serial titles. For each of these titles, we need to record which issues we have received and what the gaps are in our holdings. Back in the 1980s this information was recorded in what is called a "visible index". A visible index comprises bays of cabinets, each containing thousands of index cards. In 1985 we embarked on a project to convert all of these hard copy records to a set of records that are accessible online. The project was completed earlier this year, after 22 years of effort. The result is that this information is now much more accessible to our readers.

Just as the task of collecting has taken on new aspects which Sir John could not have anticipated, the same applies to the task of bibliography. Bibliography is now undertaken by libraries as a collaborative online activity. And the products of bibliography are put to work to help make the process of finding and getting information as rapid and easy as possible.

To a very large extent that “collaborative online activity” by Australian libraries takes place within a service called Libraries Australia, provided by the National Library but managed for and on the advice of the Australian library community.

For the past 26 years the libraries of Australia have been collaborating to share their cataloguing information and to build what we call the national bibliographic database. Since February last year this database has been available worldwide to the public through the free online service that is part of Libraries Australia. The database records over 17 million different titles and editions, published in Australia and overseas, and held in more than 42 million copies by over 800 Australian libraries. A simple search now enables anyone to verify a title, locate a copy, and to put themselves on the pathway to obtaining or consulting that copy.

Large and important as that database is, our aspirations now encompass an even more extensive space. We are planning to compile a wider national data store which will include information about pictures, research papers, journal articles and tens of millions of newspaper articles.

This means that a single search on (say) the name “percy grainger” will be able to retrieve hundreds of books, pictures, music scores, sound recordings, manuscript collections and newspaper articles associated with Percy Grainger.

We are also conscious that young, and not so young, people most frequently start their information search in Google and are often blissfully unaware of services like Libraries Australia. So we are putting our information in their way, by seeding Google with the data held in our national data store. People who search Google will increasingly find links to books and other resources held in Australian libraries. We are, if you like, working to “weave libraries into the Web”.

## **Digitisation**

In his 1962 recording, Sir John Ferguson said that “students are often put to great inconvenience through the absence of copies in their neighbourhood”. He was speaking of historic newspapers, but his observation could apply to any of the resources collected by our libraries. This traditional limitation on the researcher is now being steadily addressed by the progress of digitisation, combined with the reach of the Web. Every year a greater proportion of these research resources become accessible in digital form.

At the National Library we have been digitising portions of our collection since 1996, with an expanded program since 2001. We have now digitised:

- around 100,000 items in our Pictures Collection, ranging from pictures depicting European exploration of the South Pacific through to contemporary events
- over 10,000 Australian music scores from the 19th and early 20th centuries
- over 6,000 maps which range from the earliest Dutch charting of Australia through to 20<sup>th</sup> century depictions such as Walter Burley Griffin's vision of Canberra
- over 8,000 manuscript items including James Cook's Endeavour Journal and the papers of Sir Edmund Barton.

The National Library's holdings of maps have been built on the foundations of some wonderful collections assembled by Petherick, Nan Kivell, Tooley and of course Ferguson. Sir John Ferguson collected maps of the Sydney region, the Hunter Valley and other places in New South Wales, including many land sale maps. If you go to Libraries Australia and enter the search "ferguson collection" limited to maps you will discover over 400 maps collected by Ferguson. All of them have been digitised. This means that you can examine every one of these maps from the comfort of your home, scrolling across them and zooming in on the smallest detail.

In his 1962 recording, Sir John Ferguson also said that "the most valuable publications for historical research will always be our early almanacs and newspapers". He would have marvelled at what is about to become possible in terms of universal access to our earliest newspapers.

The National Library has now commenced a project to digitise 3 million pages of Australian newspapers, spanning the period from 1803 to 1954, and covering newspapers from every state and territory. This project will create a text searchable database that will be freely available online. "Text-searchable" means that we aim to make it possible to search on every word in these tens of millions of newspaper articles. Someone doing a search on, say, the name "percy grainger" will be able to view an image of every newspaper article in the database in which this name appears.

Needless to say, this online service will dramatically enhance access to historic Australian newspaper articles. It is likely to have a wide audience, including social science researchers, family historians and the general public.

The newspapers will be digitised by commencing with existing microfilm versions, where these are of sufficient quality. The microfilm will be processed to create digital page images, and these page images will be further processed to create data files of text and article images following Optical Character Recognition (OCR) processing. The National Library is building a system to manage this content and to provide the search and delivery service.

The National Library is leading this project, but we are working closely with our colleagues in the state and territory libraries who are providing advice on the best newspapers to digitise. The state and territory libraries are also undertaking projects to re-microfilm newspaper issues where necessary, and many of them are also seeking resources to fund the extension of the project to additional titles, including regional newspapers.

Our immediate plans are focussing on newspapers such as the Sydney Gazette, the Argus, the Courier Mail and the Adelaide Advertiser. We will begin with the first issue of the Sydney Gazette in 1803 and continue with a range of other titles through to the end of 1954, at which point we encounter the barrier of copyright.

We will begin to offer this service in the first half of 2008. Initially it will contain only about 500,000 pages covering two or three newspaper titles, but we expect the database to grow by late 2010 to comprise about three million pages (around 20 million articles) covering eight or nine titles.

## **Conclusion**

Towards the end of the 2001, the National Council for the Centenary of Federation held one its last exhibitions, called "Peoplescape". The exhibition comprised over 5000 life size plastic cut-out figures, arranged on the lawns over and around Parliament House, and representing people who have made a major contribution to the nation. The National Library was asked to make two nominations for Peoplescape. We made one for Pauline Fanning, who worked at the Library for 44 years until 1980, and made a major impact on the development of our Australian collections. Our other nomination was for Sir John Ferguson.

Sir John Ferguson died in 1969 without seeing the newly opened National Library building that now housed his collection of 34,000 volumes. But his dedication to collecting and bibliography will never be forgotten. Our understanding of the history of Australia will always be enhanced by his dedication to that same cause.