Keeping the culture: archiving and the 21st Century

Dr Bob Pymm
Manager, Collection Management
ScreenSound Australia
(formerly the National Film and Sound Archive)
bob_pymm@screensound.gov.au

Abstract:
Archiving for permanent retention is facing some major challenges as we move into the next millennium. These include issues relating to selection from a burgeoning mass of information being produced in a wide range of formats; issues relating to media longevity and equipment obsolescence; migrating information across formats; the commercialisation of activities; the growing impact of IT requirements and the complexity of copyright and other rights in digital materials.

Firstly, what comprises the culture we want to keep? Artefacts, print materials, electronic records and a broad range of film and sound materials, both analogue and digital, of which only a fraction are ever kept as cultural records for informing future generations about our way of life. The whole question of selection decisions is a huge one, crucially important and which will be touched upon later in this paper. Just consider the audio-visual industries with which my organisation is chiefly concerned. Twenty fours a day, seven days a week, 365 days a year, over 50 commercial television stations, Skychannel, Fox and cable TV, as well as the ABC, SBS and infant community stations are broadcasting television around Australia. At the same time, hundreds of radio stations, national, commercial and community are broadcasting sound within Australia. In addition, over a hundred film distributors are releasing around 300 feature films a year and about 120 video titles a month. And finally, in any one year, an estimated 5500 rock bands perform within Australia (goodness knows how many performances), 1.5 million people attend a live performance of popular music, 1200 Australian music scores are written, over 60 orchestras and 250 choirs perform in front of around 300,000 people and over $300 million is spent on sales of recorded music (The Australian Cultural Industry: Available Data and Source, 1990). And this was in 1990 (the last time the Department collated these figures) – no doubt it is even more today and will continue to grow rapidly (as an example of this expansion, the Library of Congress currently hold around 3 million audio and video items which they estimate will grow to around 22 million over the next fifty years (Storm, 1998) – a fairly daunting prospect!).

With print sources still expanding (over 10,000 books published annually in Australia and hundreds of newspapers and magazines), and the burgeoning growth of the digital medium (estimated in mid-1997 at two and a half terabytes of information (Lyman and Kahle, 1998) – probably double that now), together with photographs, manuscripts, sheet music publishing etc etc makes the task of maintaining anything other than a tiny proportion of this output in any of the country's keeping institutions a major challenge.

From these statistics it seems that while there are a large number of technological issues that need to be addressed and resolved in a way that provides for long term security of collections (and these will be considered later in this paper), perhaps one of the biggest issues facing institutions as we move into the third millennium, is how to make selection decisions from this enormous range of materials that will ensure the relevance of our collections to future users and provide the appropriate finding aids and delivery mechanisms to make this material available. So before we get into consideration of changing technologies, their permanence or lack thereof and the problems the pace of technological change causes, lets first of all consider this selection issue.

Selecting for the 21st Century Archive

We no longer share the confidence of those who established our major research libraries of a hundred years ago. Described by American, Dee Garrison as 'Apostles of Culture', these predominantly, male, white, middle-class, liberally educated elites had few doubts as to what they should or shouldn't have in their libraries. As Smith, an American academic observed, "those with cultural power… tend to be members of socially, economically and politically established classes (or to serve them and identify their interests with theirs), the texts that survive will tend to be those that appear to reflect and reinforce established ideologies" (quoted in Cyzyk, 1993). Thus being products of their place and time with all the cultural baggage this entails, through no fault of their own, huge areas of what would now be

regarded as important, interesting and revealing were neglected. Our late nineteenth century gentlemen would have been unlikely to regard material emanating from factory workers, domestic staff or prostitutes as worthy of a place in his collections - yet these groups played an important role in Victorian life and more primary sources from them would be invaluable today.

Now, with our appreciation of a multicultural and pluralistic society, we are less confident in our assertions over selection decisions. Despite carefully developed policies, collecting and particularly, archiving for all time, is still, to a degree, a subjective process influenced by many factors. Thus articulate and high profile pressure groups may make themselves more visible in society and be more proactive in lobbying institutions to build relevant collections, while less vocal, but just as relevant organisations may be overlooked. Groups such as Greenpeace and the Gay and Monarchist and Republican movements are publicity conscious and highly professional in their approach to the media and in their understanding of the importance of archiving their activities. Other groups, with equally valid concerns, may be less successful at this and in order to maintain a balanced approach to collecting, institutions may need to be proactive and go searching for relevant material (examples from my own organisation are the Country Women's Association - illustrating changes in rural Australia, and the experience of illegal migrants in detention centres in NW Australia – neither are vocal or pushy yet have played an influential role in the development and changes in our society).   Given the vast and rapidly increasing amount of material out there, and the pressure on our limited resources, undertaking proactive collecting programs can sometimes fall by the wayside with the result that important parts of the community and its cultural output may not be properly represented in collections.

Selection decisions are becoming harder - no doubt about it. Public and political scrutiny is growing (would anyone promote the development of a collection of Nazi and related material? Of pornography?) and funding will always seem to be insufficient for what needs to be done.   Owners of material are far more aware of its potential 'value' and, particularly for audio-visual, pictorial or digital/electronic material, are reluctant to hand over any rights in the material, making the policing of access and copyright control an increasingly demanding part of providing public access to collections. And, if it is seen that money can be made from providing access or promoting a collection, then it is likely that commercial providers may move into our 'traditional' area, competing in the marketplace. Again, an example drawn from my own institution relates to cinema newsreels that ran in Australia from the 1920s to the mid-1970s. ScreenSound holds a major collection of these materials but all copyright in them is held by Filmworld Australia. Thus clients may view and order copies from us, but first get permission and pay Filmworld for the rights. This use of a middle-man is something of a concern, offering governments driven by the desire for cost savings and privatisation, ready opportunities to outsource activities traditionally associated with our institutions.

Thus the massive increase in output, and the different mediums and technologies on which it appears and is supported leave us in the position today of having to be even more selective on what is to be represented in our cultural archives and having to take into account many more factors than just the content - playback technologies, accessing software, availability of hardware, long term preservation issues, supporting documentation and difficult copyright and access issues all make the final selection decision ever more complex and demanding.

The medium and the playback

As Professor Lehmann, a keynote speaker at the 1996 VALA Conference, noted:

"...the new knowledge (electronic information) in its most prevalent form is short-lived, transitory, user-oriented knowledge - the right information with the right software at the right time. While information is theoretically indestructible, the medium that carries it is actually quite fragile."

Unlike print based collections where there is really only one area of preservation concern, namely the item itself, most of the newer mediums are not intrinsically readable without some form of intermediary, ie. a playback mechanism. Thus preserving and accessing collections into the twenty-first century is going to require far more attention to preservation issues which will be more complex than those that exist for print and paper materials. Even maintaining access to the information stored in the new mediums will be difficult enough, let alone the problem of maintaining the physical item itself and the associated playback hardware and software.

It would seem that preserving the physical item itself may be more straightforward (this is sometimes known as 'product persistence'), eg. Kodak claim with confidence that writeable CDs will last 100 years under office or home conditions, 200 years if stored in archival conditions (cool, dry and dark). However, it is hard to imagine that CD players as we know them will exist in 30 years time, who knows about 200. So while the medium itself may last the distance, the hardware or software to actually use them is unlikely to have such longevity. This is in contrast to the mechanical technologies in use in the first half of the century where gramophones and film projectors can be easily preserved, readily maintained and if necessary completely reconstructed by skilled technicians. This will not be the case for modern equipment relying upon printed circuits, specially designed microchips and in-built software.

Thus the need to 'migrate' data from one medium to another at regular intervals. This may be quite simple such as copying computer files using compatible software and writing to a different medium. However, it can also be incredibly complex and costly requiring specialised hardware, software, storage mediums and time consuming checking and error correction processes to ensure good quality copies are made. If this has to be done every ten or twenty years, it could tie up a significant level of archival resources. And of course, how often should it be done? With print and paper materials, deterioration is quite readily ascertained from visual inspection by an experienced staff member. For audio-visual or electronic materials, checking for deterioration or chemical instability is far harder.

Risk management programs, where samples of the collection are regularly pulled out for detailed (and again, time-consuming) inspection become increasingly important. Thus equipment and other resources are needed to 'inspect' the collection and measure deterioration as this will be one of the major variables to determine if migration is necessary. This resource intensive work is undertaken on an on-going basis at ScreenSound with a recent project, involving the examination of 300 ten inch audio tapes of a particular brand and formulation (from a collection of over 10,000), requiring staff to physically examine each item to identify extremely fragile tapes, audition those that appeared to be physically OK, respool these and update their entries on the collection management database. Tapes identified as extremely fragile or with a noticeable loss of signal (33 out of 300 or 11%) were passed to expert conservators for individual treatment before copying. The process took 96

staff working days and while it can be assumed that the rest of the collection has a similar level of degradation, resources are not available to continue through the entire 10,000. The risk involved, with say 11% of the collection in possible danger, has to be weighed against competing demands – there is no easy answer.

Our next project is to survey the colour film collection looking specifically for changes in the colour balance which we know, is the major problem facing modern colour film. Traditionally, this work has relied upon a densitometer and interpretation by a trained user – a fairly subjective process unless significant degradation is present. With digitisation using sophisticated equipment, it is possible to encode at extremely high resolutions a particular film frame. The software will then store the red, blue and green components of the image as a number based upon the levels measured by the digitisation process. In say, five years time, we will be able to repeat the process on the same frame of film and readily compare the results and assess the changes.   This is a far more accurate and sensitive process than has been possible before and should lead to greater levels of understanding of the deterioration process.

Back to the migration process and a commonly accepted model is to try and break down the effort required into components. Thus there will be a basic hardware consideration – do we still have a machine that can read eight inch floppy disks?   At the next level there are operating systems – Unix, DOS, CPM etc – how long will they be supported, how backward compatible are new releases?   Moving up is the development environment – is our application developed using a proprietary product such as Oracle, Lotus Notes etc. and how long will it be around?   Finally there is the program itself and the data it accesses – an example is ScreenSound's collection management database, MAVIS.   This was redeveloped at significant cost a couple of years ago to run in a Windows environment – no doubt a similar process will have to be repeated at fairly regular intervals (although products can last a reasonable length of time – look at ABN!). All of these components have to exist and be compatible to enable materials to be copied, migrated to new media or new software platforms or remain accessible to users.

Videotape provides a good example of the complexity of this issue at the basic, physical level. Currently available formats include the ubiquitous VHS, Super VHS, 8mm, 3/4 inch, Betacam and 1 inch C format - these are analogue tapes. The digital market has seen a rapid increase in the number of proprietary formats in recent years including Digital Betacam, Panasonic's C-1, D-2 and D-5 formats. With the digital formats there is the added concern, particularly for archives, of the risk of image loss caused by bit-rate reduction where various different compression techniques are being used. Even with digital TV, Australia and Europe are going to run on different systems to the US – a repeat of the PAL/NTSC situation with analogue transmissions.

Most of these video formats need their own playback machine with appropriate software to be readable. Copying these tapes from one format to another is an expensive business both in hardware requirements and the staff time necessary to undertake the work. This is not a new phenomena, with earlier formats such as two inch tape, which can only be read on large, expensive playback machines which ceased being made twenty years ago and are very difficult to find today. ScreenSound has a large collection of 60's and 70's popular TV on two inch such as Homicide, No. 96, The Box etc. Access to these, much of which has not been given priority for copying, is difficult, time-consuming and thus expensive. The life of our

two inch machines is limited and it maybe that it will not be possible to copy everything raising the thorny issue of selection once again.

Frighteningly, the pace of format obsolescence is increasing all the time - it is difficult to pick the longer term survivors from the 'here today, gone tomorrow' formats - yet this is a crucial task for anyone involved in the selection decision.   My deputy director, Ray Edmondson, talks of the *tyranny of technology* for audiovisual archives with them becoming technological museums desperately trying to keep old formats alive. And this isn't only true for audio-visual collections with libraries now facing the same difficulties regarding computer files and electronic publishing and traditional archives having to cope with the electronic records created by parent organisations using regularly changing software and hardware. Thus not only will there be the tyranny of technology but we will also be at the mercy of the manufacturer, who may well upgrade or discontinue hardware or software products or change platforms with little concern over the impact this may have on long term access and preservation. Will all this cause the emphasis of our organisations to change, becoming dominated by preservation, IT and technical activities (and the consequent budgetary demands these make), rather than predominantly with content?

To conclude, a 1997 report to the Library of Congress prepared by William Nugent, talked of "the inexorable degradation of all current information carriers", of the "mortality of information" and summarised the situation thus:

> To achieve preservation in the face of the inexorable decay of all information carriers, the information must be copied before the medium deteriorates.
>
> Analog copies cannot serve as they suffer severe losses generation-to-generation.
>
> Digital copies, protected by error detection and correctional methodologies, retain their integrity over an indefinitely large number of generations.
>
> It is therefore imperative to begin digitization of those library collections at greatest risk of impending decay.

The trouble is, will we have the resources to undertake the work at a sufficient level to effectively handle existing collections and cope with the future? The strategy seems clear, successfully implementing it will be a major challenge. However, there are a number of major international projects, currently underway, that are addressing these issues by funding projects and researching methodologies to simplify and standardise practices.   Thus the UNESCO Memory of the World programme, running since 1996, has supported a number of digitisation projects of manuscript, print and audio materials in a range of countries. The programme has also been instrumental in building commitment, particularly in non-Western countries, to the principle of preserving a wide range of cultural artefacts significant to a nation or its region. A report of their progress is available at http://www.unesco.org/webworld/memory.

Another international programme, the InterPARES project, hosted by the University of British Columbia, has brought together archivists, software and hardware engineers, government and private industry to address the problem of the permanent preservation of electronic records. They have adopted the stance that "technology is expected to evolve into the foreseeable future, therefore, theories and methodologies for the long-term preservation

of electronic records cannot be dependent on specific technologies but must be centered on the nature and meaning of the record itself". (http://www.interpares.org/background). Thus not only is this project looking at technical matters it is also coming to grips with basic definitional issues which may impact upon the very nature in which information, and the relationship between informational elements, is stored. This is an important project and one to follow with interest, despite what appears to be the deliberately complex and obscure language used in their reporting! (an example…. "To develop the Template for Analysis a form of grounded theory is being used. "Grounded theory" is a method for discovering concepts and hypotheses and developing theory directly from the data under observation.")

Commercialisation of activities

As many of you will be aware, this is not a new challenge. I remember hearing Maurice B Line from the British Library talking in the early 1980s of how the BL had been told by Margaret Thatcher it had eight years to move from an overwhelmingly government funded organisation to one where 50% of its income was raised through its business activities. Apparently they managed it with a year to spare. Many of us are facing similar strictures today and this will be a major challenge for the coming decade.

At the same time that we are being 'encouraged' to raise our level of self-sufficiency through various income generating activities, it is becoming increasingly easy for competitors to enter our market and directly challenge the monopoly situation most of us have enjoyed in the past. To use economic talk, the barriers to entry into the archiving and preservation market have fallen significantly with the changes being wrought by technology. Thus it would be prohibitively costly for any commercial enterprise to build a huge print-based collection to compete with our established research libraries, yet comparatively easy and cheap for anyone with the necessary expertise to provide on-line access to databases, electronic publications and related materials and expertise. As more and more material appears in a digital form (including graphics, pictorial, sound and moving image) and the technology to distribute, download or otherwise provide access becomes cheaper and more pervasive, the more likely it is that commercial enterprises will move into some of our traditional markets. Already, the established commercial players such as Silverplatter, Dialog, UMI, Elsevier etc act as archives providing huge volumes of information. But they do this as part of a business enterprise, not to support the cultural heritage of a time or place – do we need to be concerned about this?

Inevitably, this growth of commercial activity will require established organisations to become more business-like in their approach to preservation and access. The development of partnerships between traditional institutions and some of the new, technology-based players will slowly but almost certainly increase. We both have skills and resources to offer that complement each other and offer benefits to both the organisations and their clients. Another inevitability is the increasing costs that will be placed on users in order to gain access to materials and in particular, to copy or download information. With the vast amount of material available, the range of formats on which it will be stored, the need for regular copying or migrating of data (and the digitisation of older materials), the need for specialist software and hardware to access and copy, tighter control over copyright and other ownership related rights and the increased demand from users for on-line access to the item will all bring a significant cost that cannot be born by the institution alone. These, together with the need to cover costs and quite possibly, contribute directly to the running of the organisation, will

mean a gradual increase in charging and the level of those charges. It is not clear how this growth of user pays will impact upon traditional users and the profession will have to carefully juggle the transition that is currently underway and unlikely to be reversed.

Copying, access and ethics

Once materials are available digitally, the problem of control over access, manipulation and exploitation becomes a major one. Users frequently are not even in our institutions when they are searching the catalogue or gaining access to digital information – controlling their use of it is virtually impossible. It is quite straightforward for institutions to get permission for images to be digitised and placed in on-line catalogues but how do you reassure copyright owners that their images will not be downloaded and used for purposes they may not agree with? And how do we protect ourselves if they do? When a recent proposal was put forward by a commercial organisation to rent or sell full length features, downloadable from the Net, a Los Angeles attorney was quoted as saying "the amounts of money at stake are so huge……studios do not want to give away their after market revenue. If I'm Disney, I really don't want to have 100,000 perfectly reproduced digital copies of my films flying around the world." A reasonable concern!

The popular press frequently carries stories of court cases concerning the unauthorised use of individual images that have been digitised, enhanced and used out of context to promote products (Fred Astaire promoting condoms was a recent example that his heirs brought to court). The integration of actual newsreel material in popular films (for example Forrest Gump), 'bringing back to life' artists long dead and creating images (static or moving) that never existed in actuality are relatively easy to achieve but raise huge difficulties regarding rights, ethics and responsibilities, even if legally, it is acceptable.

With this ability to alter and manipulate data arises the possibility of 'altering history' – in our efforts to copy or clean-up a text or an image it is possible to actually create something that never existed and was never intended. Governments can sometimes be very keen on 'changing history' and the ability to actually falsify events is now readily available to them.

None of this is new, it existed in the analogue domain. It is just that now, it is so much easier to achieve, so much harder to prevent and quite possibly, so much more commercially attractive.

So, what are the challenges we face in the relatively near future? I think these can be summarised thus:

- Selection decisions will become more demanding. Selection not only for acquisition of new material but also for what will be migrated and onto what format. Professionals involved in this work will require not only an understanding of collection development and its application in their organisation but also a good understanding of technical formats and their strengths and weaknesses for long term preservation.
- Inevitably, collection development policies will become more complex. They will be revised more frequently; be developed in close cooperation with kindred institutions and input from external interest groups will be fundamental.
- Acquiring the rights to provide access to certain types of material will become more complex, contentious and thus costly. Conversely, access may not be granted at all with

a greater proportion of archival collections being 'on deposit' only with the assumption that at some future date, rights to this material will revert to the institution. This is quite common practice today with commercial film makers, where ScreenSound holds film prints on the understanding that we will either acquire rights 50 years on or will levy commercial storage charges for the period.

- A smaller proportion of 'published' materials will end up residing in government funded institutions. This has a number of implications relating to access and availability with increased charges and restrictions a seeming inevitable consequence. Increasingly, access clearances will take more time and require far more stringent policing and record keeping.

- Commercial organisations handling the distribution of digital material will grow in importance due to their ability to concentrate on particular areas of specialisation, focus on target markets and the relatively low barriers to entry into the market.

- The growth of partnerships between commercial and public institutions will be an inevitable and practical response to these changes. Once initial teething troubles have been overcome this should lead to both parties gaining through the exchange of ideas, corporate knowledge, corporate assets and, importantly for public institutions, capitalising upon their established prestige and goodwill.

- This mix of archival repositories will probably lead to charging policies that will reflect the true cost of providing a service, possibly to the detriment of some traditional client bases.

- Accountability, transparency and full cost recovery will become ever more important as governments move to limit funding growth to traditional keeping institutions.

- The number of obsolete mediums to be stored, copied and preserved will increase enormously, as well the associated hardware and software necessary to undertake this process. This will lead an increasing percentage of budgets being spent on preservation and associated tasks, including IT infrastructure.

- An increasing amount of organisational 'energy' will be devoted to copying older materials onto newer formats, managing these new items, describing them sufficiently in catalogues or databases and making decisions as to the fate of the superseded formats – disposal is never straightforward.

- The need for increased IT and other technical expertise may lead to significant changes in the balance of staffing and thus corporate ethos in many organisations. Major decisions may well be made based o technical rather than broader, sociological considerations.

- The digitisation of existing collections will be a slow, on-going process that will concentrate on materials out of copyright control or small, specialist collections. Tackling other materials will be far more difficult and costly with all established institutions having to maintain traditional collections for many years to come.

A somewhat daunting list and no doubt incomplete. But these changes also offer an exciting and demanding future where, like virtually every other profession, we will have to evolve to meet the challenge. It may be happening rather more quickly than many of us would like, and some of its impact will fundamentally change our role in society, but this is inevitable and has happened in the past - no doubt we'll somehow keep up!

Bibliography

Cyzyk, M.   Canon formation, library collections and the dilemma of collection development, *College and Research Libraries,* Jan 1993.

Garrison, Dee.   *Apostles of culture: the public librarian and American society 1876-1920*, New York: The Free Press, 1979.

Lehmann, Klaus Dieter. *Preserving digital incunabula*, Melbourne: VALA Conference, 1996.

Lyman, Peter, Kahle, Brewster.   Archiving digital cultural artefacts, *D-Lib Magazine,* July/August 1998.

Nugent, William R. *Digitizing Library Collections for Preservation and Archiving* (Preservation Research and Testing Series No. 9705), Washington, Library of Congress, 1997.

Storm, William D.   *Unified strategy for the preservation of audio and video materials.* (Preservation Research and Testing Series No. 9806), Washington, Library of Congress, 1998.