Submission to the Joint Committee on Publications

Answers to supplementary questions—National Library of Australia

Hard-and software

1. What steps are you taking to maintain accessibility of machine-readable material?

"Machine-readable material" can be taken to cover a wide range of materials. In terms of the National Library's collections it can be divided into 3 broad categories:

i. Digital materials, which may be held on discrete physical format carriers such as floppy disks, CDs, digital tapes, or in computer storage – these are of greatest concern to the Library. They are subject to loss through changes in access technology, through failure of the carrier, and through corruption of the data itself.

ii. Non-digital electronic materials such as analogue sound and videotapes, and gramophone records. These are of significant concern. They are subject to loss through the same influences at digital materials.

iii. Non-electronic machine-readable materials such as microfilm and movie film. By some definitions these are not counted as "machine-readable" as they can be accessed without using anything more than a magnifier, even though they are normally accessed via some kind of viewing machinery. They are most subject to loss through damage to the carrier.

i. Digital materials:

The steps we are taking to maintain accessibility fall into a number of categories:

creation –

- the use of standardised open formats when we are creating digital objects, able to be migrated
- use of suitable quality physical format carriers where necessary for storage

documentation –

- use of persistent identifiers (under investigation) in naming files
- adequate documentation to identify material and make it findable
- adequate metadata to support long-term management and preservation
- possible identification of the "essence" needing to survive any preservation action (under investigation)

management –

- allocation of responsibilities and resources, development of policies and procedures
- research to support decision-making, including keeping up to date on developments elsewhere
- developing appropriate technical infrastructure
- knowing what is in the collections
- identifying risks

- identifying priorities

storage –

- storing physical carriers in suitable conditions to minimise rate of deterioration
- disaster preparedness
- separating access copies from preservation masters
- maintaining equipment needed for access and for preservation processes
- maintaining software
- identifying suitable indicators for preservation action (under investigation)
- testing the condition of carriers

preservation actions –

- back up copies, especially on different carriers
- refreshing media and transferring from less stable to more stable media
- investigating procedures for migration, emulation and other preservation pathways
- migrating material (under investigation)
- finding suitable emulation software (under investigation)
- recovering data from inaccessible formats (under investigation)

For **non-digital machine-readable materials**, most of these steps still apply. We need to pay more attention to the longevity of the carrier, as the media refreshing cycles tend to be much longer, and we sometimes need to undertake repair of damaged carriers.

2. In the NLA's view, how long is it reasonable to maintain the machines and software on which some of your material depend?

The Library's underlying intention is not to rely on machines to maintain accessibility, simply because of the difficulty and cost of doing so – as time goes by it will become expensive, and then probably impossible, to maintain the parts and expertise needed to keep them operating. Having said that, we realise that for some materials we will have to maintain equipment as a bridge to more effective preservation pathways.

Even in those circumstances we will aim to minimise that period in which we need to retain obsolete equipment.

The Library does not believe it will be feasible to maintain complicated computer equipment for more than a few years past the point at which general systems support has been withdrawn.

If we adopt migration as a pathway, we will need to take account of how long the migration process will take, and migrations will generally have to begin well in advance.

Some basic access hardware (such as disk drives of a particular size) will be kept longer as one-offs for specialised preservation processing.

An extreme example of this is the analogue audio playback equipment that the Library will need to maintain for its analogue-to-digital conversion, which is expected to take around 20 years. Fortunately, the analogue equipment is robust and relatively easy to maintain, and we

expect to be able to service it for such a period. Such a timeframe would not be available for computer equipment.

Software will be treated differently. The Library is investigating its options for maintaining access to an archive of software that will support the 'native' formats of its digital collections, in case migration is unsuccessful. We recognise that this is likely to be complex, given that most software depends on other software (such as operating systems) that may be linked to a particular set of hardware. However, at this stage, we see no alternative.

3. What is your policy on migration of material and what do you see as the dangers of migration?

The Library has not finally decided the extent to which it will rely on migration to overcome changes in hardware and software. At this stage we assume that migration will be a major part of our strategy, and we are actively investigating it and planning for it. However, there are a number of problems we can foresee, and we are interested in other approaches that have been suggested, especially the use of emulation software that would allow digital files to be accessed as if they were running with their original software and hardware. This approach is also problematical and unproven.

Many parts of the Library's digital collections have been created in, or already migrated to, ubiquitous standardised open formats that should be relatively straightforward to migrate to replacement formats. Examples are image files written in TIFF format, and digital audio files written as uncompressed linear data files to CD-R standards. Migration will require conversion tools that are likely to be available. Given that such collections are large and homogenous, we expect to be able to apply the conversion tools automatically.

On the other hand, with more complex objects, or with less standardised formats, migration becomes more of a problem. Many formats coming into the collections may not migrate without losing some functionality. Because different programs operate in different ways, these losses are likely to compound with each migration.

Even if these effects can be overcome with programming intervention, to do so may well be too costly for general use. Such an approach may also complicate later migrations by introducing new non-standard elements into the digital object.

There is also the challenge of migrating the different formats (text, sound, video, virtual reality, etc) in multimedia products, at different times and to different new formats, and reassembling them after migration.

None of these problems are insurmountable, but as many commentators have noted, it looks like an expensive and "heroic" effort that has to happen repeatedly. We believe it will be possible to automate some of the processes, but it will still come at a considerable management cost.

The Library is currently looking at the file formats in its digital collections and identifying formats that should be easy to manage by migration, and formats likely to cause problems. We will then assess the tools and management approaches we could use to deal with them.

Over the longer term, many people have also pointed out the dangers of relying on a strategy that could easily fail because of changes in priorities, unavailability of resources, or even just management failure. The window of opportunity for each migration is likely to be fairly narrow, so a failure of will or resources, even over a period of a few years, could jeopardise the whole strategy.

There are a number of projects overseas that are doing similar practical research on emulation approaches. Practical experience with both – and other – approaches will help us make more informed decisions about the best mix of strategies to use and how best to manage them.

Preserving and Accessing Networked Documentary Resources (PANDORA) archive

4 In your oral evidence you state that Sweden and the Netherlands have been active in preserving digital publications. Would you outline what you have learnt from the national libraries methods?

There are active digital archiving and preservation programs at various stages of progress in Canada, Germany, Finland, Norway, the UK and the USA, and other countries as well as Sweden and the Netherlands as mentioned in the Library's oral evidence. Most of these projects are exploring approaches that are influenced by their own context, but many of the projects are in close consultation with each other.

The national libraries of Sweden and the Netherlands have taken approaches to the archiving of digital publications that differ significantly from that adopted by the National Library of Australia. This has provided the Library with a useful benchmark for its own archiving and digital preservation work.

The National Library of Sweden takes regular snapshots of the entire Swedish Internet domain (ie. all sites with the address *.se*) using a software robot. This information is stored on tapes. The National Library of Australia on the other hand is very selective about what it archives and uses a software robot to capture publications in the Australian domain on an individual title by title basis. From our observation of the Swedish model we have been convinced that our selective approach will lead to a more satisfactory long-term outcome. The Swedish model of comprehensive capture and archiving has thrown up major problems. For example, because they capture a massive amount of information at once, they can not tell if they have actually captured *.se* Internet sites which are intact and useable. In addition, they have no satisfactory way of providing access to individual sites and they have not sought permission from the copyright owners of the sites to provide public access to the captured sites.

The Royal Library of the Netherlands is a leading partner in a project called NEDLIB, which is developing a prototype system for archiving and preserving digital publications. The NLA has been in regular contact with the Royal Library concerning developments on NEDLIB. Of particular interest to the NLA is NEDLIB's use of the OAIS model (Reference Model for an Open Archival Information System developed by the Consultative Committee for Space Data Systems) as a basis for the design of their prototype archive. The OAIS model is a functional specification for a digital archive. The NLA has also reviewed the OAIS model and incorporated elements into its own functional specifications for a digital archiving system.

Some of the issues on which we have been sharing information are:

- Selection of material to be archived: Some projects have decided to try to collect comprehensively, while others, including NLA, are much more selective. This is usually based on a pragmatic decision about how far the available resources can stretch. We have given priority to being able to manage and provide access to what we collect, whereas some others have decided to focus on collecting, which they acknowledge will build in very large access and management problems to be dealt with later. There are also differences in the complexity of what is being collected. While most programs including NLA's, are trying to capture material across the full range of formats and genres, some such as the National Library of Canada have restricted their intake to the most simple

formats, often text documents that are also available in print. This is attractive as a way of reducing the complexity of the task, but they do acknowledge that it not giving them a good foundation for dealing with more complex material in a production-scale archive.

- Relations with publishers and rights owners: This is a major issue for all programs. Some national libraries operate in a context of legal deposit legislation for digital publications, while others do not. In both cases programs have found there are difficult issues to resolve with publishers and rights owners. Some programs, such as the European Networked Electronic Deposit Library (NEDLIB) and the National Library of Canada, have set up mechanisms to explore these issues formally with publishers, and we are learning from their experience. This is especially important for us as we have had limited opportunities to work with large-scale commercial publishers on the Internet in Australia.
- Documentation and metadata: All programs agree that adequate documentation will be a key to maintaining accessibility. The Library is part of an international group of projects that is developing and sharing approaches to preservation metadata – the supporting information that will be needed to manage preservation.
- File naming: the Library is part of a formal international working group looking for ways of giving files persistent identifiers that will make them findable whatever their location.
- Archiving models: we have been particularly interested in understanding and commenting on the Open Archival Information System (OAIS) Reference Model for digital archives that has been developed by the Space Data community. The Reference Model is being used as a guide by a number of overseas projects, and is likely to be adopted as an ISO standard. Consultations have helped us see to what extent we can usefully follow the OAIS model.
- Technical infrastructure: we have shared background documents and discussed technical issues and solutions with a number of other national libraries that are trying to deal with similar problems.
- Preservation pathways: we are actively engaged in discussing approaches for addressing longer-term preservation issues, especially hardware and software dependency. As discussed above, we are part of an international network of projects doing practical experimentation with approaches such as migration, emulation, format standardisation, and data recovery.
- Organisational issues: there is widespread agreement that managing the processes is at least as important as finding technical solutions. We share information on management approaches, costs, policies, procedures and organisational structures, including opportunities for collaboration.

5. Which federal agency is responsible for capturing and thus preserving the digitised information authored by government and parliamentary agencies in Australia?

The National Library of Australia, under the terms of the *National Library Act 1960*, is responsible for preserving Australian publications that emanate from both the government and non-government sectors. This includes publications in digital format. Individual state libraries also have this responsibility in relation to publications published within their state.

Due to the cost and complexity of collecting and preserving digital publications the Library acquires a printed copy of a publication for preservation in those cases where both print and digital versions of the same title exist. The joint National Library of Australia and ScreenSound Australia (formerly National Film and Sound Archive) submission to the

Copyright Law Review Committee argues for the extension of legal deposit to digital publications. It also argues that both organisations have the right to be selective in relation to titles they collect and preserve.

The National Library is working with the state libraries on a strategy involving the National Library and each state library in defining its area of responsibility for digital collecting and archiving. It is proposed that the National Library and the state libraries avoid duplication in the area of digital collecting and preservation, in order to cover a much broader range of Internet publications than a single institution could manage on its own. As part of this cooperation the National Library will collect and preserve Commonwealth Government publications that exist in digital form only. (Note: this does not include material that appears in another format, eg print, or material which has been digitised to enhance public access).

Digital Services Project

6. Would you expand on what is involved in the Digital Services Project? For example, what you hope to achieve, the length of the project, whether it is on course etc.?

The Digital Services Project is the National Library's key strategy for ensuring that it can manage its digital collections effectively. In particular, the Library is aiming to achieve:

- a more efficient system for gathering and managing copies of Australian web sites that the Library has decided to archive for posterity;
- a more efficient system for managing digitised copies of collection items (such as pictures, maps and manuscript items) that the Library has chosen to make accessible online through the Internet;
- the ability to migrate these digital collections to new formats as the technologies and the prevailing technical formats change; and
- improved search access to these collections, including federated search access which allows the Library's digital collections to be searched in conjunction with those of other Australian libraries, museums and archives.

The Library commenced this project in mid 1998. In December 1998 it released an Information Paper, describing the aims, functions and architectural principles of the Project, and sought feedback from interested parties in Australia and overseas. During 1999 the Library issued two procurement documents, with a view to acquiring key system components and services to support the Project.

One of these procurement documents led to the acquisition, in late 1999, of the Blue Angel MetaStar Enterprise software to support search access to the digital collections. The Library is now using this software to establish a system supporting federated searching of digitised pictorial collections of a number of cultural institutions. Several other uses of the software are planned during the next 12 months.

The other procurement process, a Request for Tender for a Digital Collection Management System, did not produce any satisfactory solutions, although it did assist the Library to identify at least one product which may provide a partial solution to the Library's requirements.

The Library will now seek to achieve its goals in this area through three sub-projects:

- Collecting System Sub-project (to manage the collection of digital publications from the Web or from physical format sources, their conversion to a standard format, and the extraction of associated descriptive information);
- Object Management System Sub-project (to allow collection managers to maintain gathering schedules, access conditions, rights information and other administrative data about each digital resource); and
- Digital Object Storage and Delivery System Sub-project (to manage the storage of digital objects and associated copies, provide access to the objects via a Web interface, and collect relevant transaction data.).

The Library aims to complete these three tasks by mid 2001. The first of them will be undertaken as an in-house development project, using state-of-the-art web gathering tools already identified by the Library. The other two tasks should proceed largely through the acquisition, customisation and implementation of existing software and any associated hardware.

7. You state in your annual report that "Guidelines were developed for the preservation and management of physical format digital collections by Australian libraries." Do these guidelines have broad acceptance across the Australian library community; and was there any attempt at benchmarking with significant international libraries?

There appears to be broad acceptance of the principles behind the guidelines. The Library has not tried to negotiate formal agreement to them, nor any undertakings to implement them, as they were always intended to be descriptive of current best practice rather than prescriptive. Many libraries do not have a preservation role with such material and will leave it to the National and State libraries. Most State libraries have been unable to allocate resources to active management of these collections beyond the material required for current reference use, but they have indicated that they see the approach as one they would like to adopt.

The Library is very aware that the approach (involving more detailed documentation and identification, copying material from less stable carriers such as floppy disks to more stable carriers such as CD-R, and storing backup copies) is not a complete preservation strategy. It developed as a way of addressing our own immediate needs to prevent further loss of early digital material. It will be incorporated in longer-term strategies as they are developed. Some State libraries have indicated that they will wait for the longer-term strategies to be developed before attempting to implement the guidelines.

The approach was developed following a widespread search for solutions in national and similar libraries overseas. This showed a widely used approach of copying material from vulnerable carriers, and the need for adequate documentation. However, our approach has received a lot of interest from overseas organisations (such as the National Library of Canada, British Library, and the Research Libraries Group) as a leading approach for dealing with physical format digital publications.

Duplication

8. Your annual report states "Discussions with other collecting agencies aim to ensure that there is neither excessive duplication in collecting nor neglect in collecting essential cultural heritage material." Are discussions held to ensure that preservation standards are not reinvented by each organisation? (output 1.1)

The kind of discussions referred to in the annual report, as they relate to preservation standards and approaches, are regular and frequent, and take place through a number of channels, including:

- management level liaison meetings with other institutions, and especially ScreenSound Australia (SSA) and National Archives of Australia (NAA);
- operational level personal contacts between staff in collecting institutions in a number of informal networks, including sound preservation managers, digital preservation staff, preservation reformatting staff. These contacts are strongly encouraged and seen as critical to achieving our objectives;
- joint production projects with other institutions especially in digitisation, microfilming, and digital preservation (PANDORA);
- joint research and strategy development projects such as the National Plan for Australian Newspapers (NPLAN) and the development of a national strategy for deteriorating cellulose acetate collections;
- special focus meetings on digital preservation issues, both nationally and with individual state libraries;
- a series of NLA position papers on managing digital collections, requested by, circulated to and discussed with state libraries and others;
- comprehensive presentation of NLA approaches and thinking on the NLA website;
- the PADI website, which carries information on digital preservation approaches in various institutions and sectors in Australia and overseas;
- the associated padiforum-l public discussion list for active discussion and debate of approaches and standards;
- involvement in a range of Standards Australia committees developing standards in the IT and preservation reformatting areas;
- active staff participation in a number of relevant professional organisations where preservation standards are discussed, including the Australian Institute for the Conservation of Cultural Materials (AICCM), International Association of Sound Archivists (IASA), Australian Sound Recording Association (ASRA), Audio Engineers Society (AES);
- participation in the Commonwealth Corporate Management Forum IT Working Group and its Digitisation Sub-Group developing best practice and sharing information on digitisation standards;
- leadership of a Canberra-based preservation microfilming discussion group, leading to publication of *Guidelines for Preservation Microfilming in Australian and New Zealand Libraries* by NLA;
- proposed establishment of a national preservation managers' network to identify and progress areas where collaboration would be helpful.

9 Does the NLA have standards for the preservation of digital sound archiving and DVD? If so, how do these procedures compare to any ScreenSound Australia may have in place?

The Library has standards for sound preservation using both analogue and digital techniques. These are based on international technical standards (a member of Library staff is on the technical committee of the International Association of Sound Archivists), and a close study of the Library's own needs. The technical details can be provided if required.

The current approach involves digital recording and editing, and writing to both professional grade CD-R and analogue reel-to-reel tape to provide storage media with significantly different deterioration patterns. This is an interim solution as the Library moves in the direction of a high bit-rate mass storage system, integrated with its other digital collections. There are a number of technical and resource issues to be addressed before such a solution can be put in place.

The Library has not developed standards for dealing with DVD as that format does not appear in its collections, which consist almost wholly of non-commercial recordings of oral history and folklore.

ScreenSound Australia takes very similar approaches, with some differences that reflect the different kind of collection it holds. Because its audio collections are made up largely of commercial recordings in various formats and on various carriers, it is required to deal with many more carrier restoration problems than the Library, and to aim for higher audio replay standards. This means that SSA has decided to remain with analogue tape as a preservation master for longer than the Library, with the intention of transferring straight from analogue tape to a mass storage system at some time in the future.

The Library did not adopt this approach because we saw advantages in being able to speed up the eventual migration to a mass storage system by first migrating to digital via available CD-R technology.

These approaches have been fully discussed between the institutions.

Access

10. Your annual report mentions the Library's Disability Action Plan and a positive finding by ACROD (that access to the NLA was favourable for people with disabilities). Have you taken any steps to enhance access by disabled people to your Internet site, and to physical format electronic publications?

Our experience is that it is better to provide ergonomic facilities and help desk support for reading room users who bring their own electronic accessibility aids to the library rather than maintaining a library of such aids. Screen reading utilities are complex to use and the technology changes rapidly. The Royal Blind Society and similar agencies are better positioned than the Library to supply aids and equipment and to provide an appropriate level of training and support.

11. There is a program called BOBBY which can be run over Internet sites to test whether they assist access by those with disabilities. Do you believe this is an effective test, or is there another program that you could recommend to the Committee?

From its experience, the Library has found that we need to do more than just use Bobby to ensure accessibility. In addition to the use of such a program there is a need to work in the following areas:

- Maintain awareness of current HREOC and WAI Accessibility guidelines ; and continue application of these;

- Inclusion of accessibility as a requirement for designers (internal and external) and web application developers;
- Use of an HTML editing tool that includes a web accessibility validator (HoTMetaL Pro. 6.0). The accessibility validator also enables efficient application of accessibility features when it detects them in a document;
- Ensure staff preparing documents for the website are trained to be aware of and to implement accessibility features;
- Use of external testing resources. e.g. for one of our web-based services we called on the services of a blind individual who tested the pages of this service using their own special web access tools;
- Bobby provides a quality check and monitoring device to ensure our other strategies are working, but only checks on a per page basis, and not the whole website. Therefore, we need to check sampling of pages on the site; and check subsites as well.

PADI

12. What are the functions of the PADI web site, and why was it necessary to redevelop it?

The PADI (Preserving Access to Digital Information) website is basically an annotated directory to information on digital archiving and preservation. It was originally set up in 1997 as a result of a cross-sectoral working group that aimed to develop guidelines for digital preservation in Australia.

The Library accepted responsibility for the PADI website and maintained it during the following two years. During this time it continued to serve as a useful point of reference for Australian preservation managers, but the Library recognised that its full potential would only be realised if it was able to reflect a wider range of international views and initiatives, supported by more discussion and evaluation.

In 1999 the site was redesigned to make its internal structures work more efficiently (reducing the resources required to maintain it) and to improve its searchability for users. Further development in 2000 will allow an existing network of overseas expert users to add linked resources and evaluation, improving both the content and the spread of "ownership", while again reducing the NLA resources required for ongoing maintenance.

The aim, which is being realised, is for PADI to be recognised as the internationally pre-eminent source of information in its field.

The Library sees PADI as playing a more significant role than just a successful subject gateway. By bringing information together it provides an excellent opportunity for new ideas to be stimulated, which may lead to more satisfactory approaches than the currently available ones.

To support this, the Library recently established a new public discussion list *padiforum-l*, for active debate of digital preservation issues. The Library is also encouraging periodic critical reviews of international progress in digital preservation. The first such review, based on PADI information, was presented as a paper to the Victorian Association for Library Automation (VALA) 2000 conference in Melbourne recently. (Available online at:

http://www.vala.org.au/vala2000/2000pdf/Ber_Web.PDF)

The PADI website (and the padiforum-l archive) can be accessed at:
http://www.nla.gov.au/padi/

Deposit scheme

13. What standards do donors/authors have to meet for the NLA to accept digital and analog material?

Donors and authors do not have to meet any standards for the Library to accept their material. However, the Library is currently unable to archive all types of digital publications, in particular those based on a dynamic database structure. This however, is not a case of the Library rejecting a publication but rather of the Library being unable to collect it for technical reasons.

14. You state that online material is not covered by the deposit section of the Copyright Act – is it the NLA's view that this should change?

The Library, together with ScreenSound Australia, made a submission to the Copyright Law Review Committee urging the extension of legal deposit to digital publications. The Library believes that digital publications should be viewed as part of the nation's publishing output and should be collected and preserved as part of our documentary heritage. A copy of the submission is attached as it outlines the view of the National Library and other Commonwealth collecting institutions [not attached].

Authentication

15 You mentioned in your evidence that you take authorship on face value. Has the NLA any plans to take a more rigorous approach to authentication to ensure the future integrity of your information?

The National Library has a collection of over five million volumes. It collects some 85,000 new collection items and adds 360,000 new serial issues to its collection each year. The Library has never authenticated authorship; rather it collects published material and relies on the user to determine the value of the material before them. We are not aware of any library that authenticates the authorship of the publications it collects. Even if the Library wished to do so, the large volume of material it collects would preclude this activity.