DIGITAL PRESERVATION:  
The Research and Development Agenda in  
Australia  

Angela Kerry  
Swinburne University of Technology  

Prepared for the Digital Continuity Forum held at  
Swinburne University of Technology, 19th November, 2001
INTRODUCTION

In *Changing Trains at Wigan: Digital Preservation and the Future of Scholarship* Dr Seamus Ross wrote:

> The preservation and re-use of digital data and information forms both the cornerstone of future economic growth and development, and the foundation of the future of memory. ¹

The management of digital material has become increasingly critical as the range and volume of material existing purely in digital form grows. The imperative is social, economic and cultural. Digital materials document, as expressed by Walters and Garrett in the Research Libraries Group seminar report, society’s issues, concerns, ideas, discourse and events². The potential loss of the cultural memory of the developed world through the degradation and destruction of digital resources, archives, ephemeral and incidental material has been recognised as a serious problem that needs to be addressed urgently by those with a responsibility for long-term access to digital material. The need for preservation of this material is enormous, as digital material is forming an increasingly sizable portion of our intellectual and cultural heritage. In the past much of the material that documents cultural histories has survived through providence rather than specific archiving practices, neglect often facilitating rather than hindering preservation. We do not have this luxury with digital documents. A pro-active approach is essential to limit the possibility of what has been termed ‘the digital dark-ages’.

This paper aims to review the existing policies, procedures, programs and projects in the area of digital continuity, address the research directions in the immediate or near future and outline the areas of research in need of further development. It will also:

- Give a broad context to the papers presented at the Digital Continuity Forum at Swinburne University, November 2001;³
- Foster an understanding of the key issues and concerns, and the projects and programs in digital preservation in Australia and internationally for conference participants who have no or limited knowledge of the subject;
- And, provide a background for the digital continuity agenda for universities in Australia.

It will focus on the preservation of dynamic and static text, images, sound archives and datafiles in the context of Australian academic and government institutions and address the preservation of both digitised material - that which is a digital record of a physical object - and ‘born digital’ material.
Digital Preservation Background

The title of an article by Alan Howell - “Perfect One Day – Digital the Next” – sums up the hopes that were raised by the possibilities of digitisation for preserving a record of the fragile physical objects and the immense and urgent challenge that such digital material subsequently issued. The digital age challenges modes of data management, of global intellectual boundaries, of communication and socialisation and most importantly our relationship to information. Technically it challenges the traditional concepts of preservation, as the preservation of digital material is not about the tangible form of a physical object for which preservation processes are well established but rather deals with the digital form which can become physically and logically inaccessible with neglect. Loss of access to material can be caused by the degradation of the storage media, loss of functionality of the access devices, loss of manipulation and presentation functions, or the disfunction in the documentation chains. The challenge in the preservation of digital material is to retain the essence of the original material by maintaining its function, content and context.

Digital material offers both challenges and opportunities for the way we access and use digital resources in the future. If archiving and preservation strategies are successful, digital media offers unparalleled opportunity for providing continuing and efficient access to the wide range of digital materials – from academic journals, to sound bites, to records of government departments. This vision of an effective and wide-ranging preservation strategy is being addressed by a number of programs and projects in Australia and overseas focusing particularly on:

- Development of preservation frameworks
- Development of policies, strategies, procedures and standards
- Technological solutions
- Collaborative efforts, programs and projects
- Culture change in the way we produce and manage digital records, manage our human and financial resources
- Skills development and training
- Education and awareness building programs

Technological obsolescence has been the primary and most obvious hurdle in digital preservation. Changes in software and hardware have rendered some digital material inaccessible and therefore useless. Jones and Beagrie identified key issues in digital preservation within the library and archive context. In summary these are:

- Machine dependency – the reliance on specific hardware and software
- Technical obsolescence – the speed of changes in technology reducing the timeframe for preservation to 2-5 years as compared to up to centuries for traditional materials
- Fragility of media – the inherently unstable media, the need for suitable storage and the potential for this media to deteriorate without external signs

There is a range of technical solutions to the preservation of digital material with varying cost implications and suitabilities. These include:

- Migration, a set of organised tasks designed to periodically transfer materials from their original software/hardware configurations and technology to subsequent configurations and technologies
- Emulation, as defined by Rothenberg, preserves the original document in the original format through the creation of an emulation specification that can be run on subsequent hardware.
• Digital archaeology, a labour intensive last resort where a variety of methods is used to access material in obsolete formats.

In ‘Digital Preservation: Problems and Prospects’ Margaret Hedstrom raises one particular area for research focus for the future. She identifies the need for evaluation of how different technical approaches can be combined to best address preservation for different needs.8

While there is still considerable work to be done in the area of technical obsolescence, the key areas of need in the digital preservation agenda are in other areas. Therefore, while technical obsolescence has been a key focus area of development in preservation, this paper will focus primarily on the preservation frameworks and infrastructure in Australia, particularly examining digital preservation within the context of the legal, policy, social and organisational issues.

What is the digital preservation agenda?

The scope of this paper is framed by several key international documents:

• Avoiding Technological Quicksand: Finding a Viable Technical Foundation for Digital Preservation, 1998;
• Preservation Management of Digital Materials Workbook, 2000;
• An Open Model for an Archival Information System (OIAS), 2001;
• And, the draft Attributes of a Trusted Digital Repository. Meeting the needs of Research Resources, produced in 2001.

The ‘digital agenda’ was formalised in 1996 when Preserving Digital Information, a Report by the Task Force on Archiving of Digital Information was published. The key findings of this report were that:

• Creators, providers and owners of digital information were the first line of defence for preservation;
• Long-term preservation of digital information on a scale adequate for the demands of future research and scholarship will require a deep infrastructure capable of supporting a distributed system of digital archives;
• A critical component of the digital archiving infrastructure is the existence of a sufficient number of trusted organisations capable of storing, migrating and providing access to digital collections;
• A certification process for digital archives is needed to create a climate of trust in prospects of preserving digital information;
• Certified digital archives must have the right and duty to exercise an aggressive rescue function as a fail-safe mechanism for preserving valuable digital information in jeopardy of destruction, neglect or abandonment by its current custodian.9

The report made recommendations for pilot projects, requirements for support systems and the development of best practice standards. The concept of ‘deep infrastructure’ supporting an integrated system of digital archives set out in this document is central to much of the work that was done subsequently in this area.

Jeff Rothenberg’s Avoiding Technical Quicksand: Finding a Viable Technical Foundation for Digital Preservation explored the technical depth of the problem of long-term digital preservation and looked critically at previously proposed solutions. The central idea in the paper is that the most feasible preservation strategy is the emulation of obsolete systems that will enable obsolete software to be run.
In late 2000, key issues in digital continuity were addressed by the *Preservation Management of Digital Materials Workbook*. The problems identified included:

- The level of awareness and interest in digital preservation which, while gradually increasing was not keeping pace with the level of digital resource creation;
- Institutions that had not been involved in traditional preservation did not have a strong sense of playing a role in preserving digital materials;
- Individual researchers lacking guidance and institutional backing to feel confident in preservation processes;
- Digital resources often being created within a collaborative program and with external funding which creates a problem with the allocation of responsibilities for digital preservation.

In June 2001 the Consultative Committee for Space Data Systems of the American National Aeronautics and Space Administration produced the draft *Open Model for an Archival Information System (OAIS)*. This aimed to develop broad consensus on the requirements for an archive that would provide a permanent, or indefinite long-term, preservation of digital information. The recommendation ‘establishes a common framework of terms and concepts which comprise an Open Archival Information System (OAIS). It allows existing and future archives to be more meaningfully compared and contrasted. It provides a basis for further standardisation within an archival context and should promote greater vendor awareness of, and support of, archival requirements’.

In August 2001 the Research Libraries Group (RLG) and OCLC produced the draft *Attributes of a Trusted Digital Repository*. It frames the debate on digital preservation and proposes a framework for establishing reliable digital repositories. It builds on the concept of ‘deep infrastructure’ established in the 1996 report and it:

- Proposes a definition of a trusted digital repository;
- Identifies the primary attributes of a trusted digital repository;
- Articulates a framework for the development of a trusted digital repository;
- Identifies the responsibilities of an OAIS compliant digital repository;
- Informs the RLG/OCLC communities of other developments necessary to implement a reliable repository;
- And, provides a formal recommendation for future work.

This report is currently being reworked and is due for release in 2002.

Much of the impetus for digital preservation has come from the library and archive sectors. In Australia the National Library of Australia and the National Archives of Australia have become the leaders in preservation practices, policies and advocacy. The university sector in Australia is another major stakeholder in digital preservation. There is a trend for publishers to move from print-based journals to digital publication. Moreover, academic communications are occurring increasingly in digital formats, including online education programs, discussion groups, bulletin boards etc. Within this context there is also an investment by libraries (within and outside the academic sector) in acquiring digital resources and converting physical resources to digital to improve access and limit contact with physical objects. The rapid conversion to digital records demands a technical, organisational, legal and economic approach to developing an infrastructure comprehensive enough to ensure preservation of Australia’s digital assets in all relevant sectors. The National Library and the National Archives have developed policies and strategies for preservation of digital materials. In particular, there is *Archiving Web Resources: A*
policy for keeping records of web-based activity in the Commonwealth Government which establishes the National Archives of Australia’s policy on the management of online resources, best practice standards and preservation practices in the government sector and Archiving Web Resources: Guidelines for keeping records of web-based activity in the Commonwealth Government, the accompanying guidelines for the policy.

LEGAL ISSUES

Rights management

In 'The Future of the Book in the Digital World' Clifford Lynch raises the issue of the global marketplace in the context of rights management. He says:

> Networked information creates a globalized information marketplace. Historically, the Internet has been a world without boarders or customs checkpoints or geography. This is at odds with the very geographically based traditions of publishing, where companies obtain the rights to publish works in specific regional markets . . . . Net-based content – which can move across the globe without the inconveniences of customs . . . threatens to seriously upset some long standing business practices\(^1\).\(^2\)

He goes on to point out that the motivation for supporting regional markets is both economic and social. The social imperative is related to the idea of the text as a trafficker of ideas, ideas that can be controlled by geography, thus preventing the propagation of controversial ideas and ideals. Digital material is redefining the needs for rights management in the areas of intellectual property, copyright, commercial interests and author rights or moral rights. The digital agenda is challenging the traditional environment of physical material in which ownership, geographic restrictions and rights and responsibilities have been clearly defined. Both the opportunities and challenges of digital material come from its radically different *modus operandi*.

In Australia intellectual property rights associated with digital materials are an issue in preservation as intellectual property rights, such as copyright right, can be compromised by digital archiving. The very act of preservation for digital materials can be an infringement of both legal and moral rights. The form and functionality of digital materials is affected by content, delivery and software. This creates problems in terms of preservation as technology that has been developed to provide copyright protection also inhibits or prevents actions needed for preservation\(^3\). In addition to this the nature of digital documents means that often there is no one authoritative original document. As the notion of copying relies on the ‘original’ document in terms of copyright and intellectual property, intellectual property rights may be hard to assert. Creators and publishers are often concerned with controlling access and potential infringements of copyright. The management of intellectual property rights is linked closely with access and authenticity.

The concept of authenticity or integrity of digital material is critically important as it relates to intellectual property, copyright, access and author’s rights (or moral rights). Technical processes, such as migration and emulation, affect the nature, form, function and appearance of digital material and therefore the authenticity of this material.\(^4\) Authenticity of a document has been defined by the CEDARS Project as a document that is the same as that which a user expected based on a prior reference and integrity of documents as that which is complete and unaltered from the time of creation. It defines the process to authenticate as to validate the integrity of a digital object with respect to its original authorised creation.\(^5\) The
National Library of Australia PADI website defines it as such: ‘The authenticity of a digital object refers to the degree of confidence a user can have that the object is the same as that expected based on a prior reference or that it is what it purports to be.’ Cornell’s PRISM Project defines integrity of digital material by the criteria of preservation, reliability, interoperability (relying on standards), security, metadata.

Digital material is easy to alter. There may be many versions of the same material or it may be altered by the preservation process either deliberately or incidentally. This raises questions about when a document can be considered authentic, which version is authentic and what are the characteristics of the material that must be maintained in order to keep the integrity of the document. The National Library of Australia addresses these issues, saying ‘Aspects such as a document's functionality, its dependence on particular software and its relationship to other documents are all features which need to be considered in the establishment of its authenticity’. To a certain extent this has been addressed through the inclusion of metadata and documentation, digital watermarking for detecting unauthorised copying and digital signatures.

It is, however, what the CEDARS Project refers to as the ‘preservable essence’ of an object that still needs to be addressed comprehensively. Authentication is vitally important in preservation of digital material as it relates strongly to the rights of creators as well as to the future need for verifiability of the material. Authenticity is an area that needs more research. In particular the following need to be addressed:

- Development of a working definition of authenticity;
- Definition of an object’s significant properties and assessment of how an object’s significant properties affect the use and access of the material;
- Consideration of at what point an object is no longer considered original and how much can it be changed by the preservation process before the integrity or authenticity is compromised;
- What constitutes the original document in dynamic resources such as internet sites;
- And, what technical mechanisms need to be in place to ensure the authenticity of the archived material.

Copyright and legal deposit

Legal deposit is seen as an important step for the preservation of digital material. In Australia it would bring digital material into line with traditional publication formats. Without legal deposit preservation of digital publications is necessarily ad-hoc. In 1999 further proposals were put for the simplification of the Australian Copyright Act 1968. These proposals included proposed changes relating to legal deposit. The proposed changes have not yet been taken forward. The Copyright Law Review Committee regarded the legal deposit provision as important for the preservation of materials that are part of the nation’s cultural heritage. They therefore made recommendations for digital publications to be treated as physical materials for the purposes of legal deposit and named the National Library of Australia and National Film and Sound Archive as repository institutions.

Legal deposit of digital publications is seen by the National Library as important for preservation as it would parallel the collection and maintenance of the physical collections. They are currently considering the possibility of legislating to harvest whole domains, an approach being trialed by the French. An issue that needs to be resolved when harvesting whole web domains is that of physical access to the material. While harvesting the entire Australian domain is technically feasible, providing meaningful access to the data still needs to be resolved.
In the absence of legal deposit for digital material at a national level alternative solutions must be considered for the preservation of the intellectual and cultural heritage of Australia. The National Library is currently developing an agreement with the Australian Publishers Association (APA) setting out the rights and responsibilities of the Library and publishers.  

Access

Access is defined by Beagrie and Jones as meaning the ‘continued, ongoing usability of a digital resource, retaining all qualities of authenticity, accuracy and functionality deemed to be essential for the purposes the digital material was created and/or acquired for’. Access is a legal issue (protecting the rights of the creator and rights owner), a technical issue (the maintenance of digital material in a way that is usable and in an authentic form) and an intellectual issue (the ability for the material to be interpreted by current and future users). Access is not just a physical or legal issue. It must encompass intellectual access to digital material, ie the ability of the user to interpret the data. It is one of the preconditions of an OAIS Model archive that the information is independently understandable. This then links to the concept of authenticity and integrity of a digital object. For digital material, particular text and images, the context, form and appearance of the material can hold meaning that dictates its interpretation. In digital preservation access involves a range of processes that include the decisions about what is preserved, rights management of the material and the technical processes that facilitate access to material, documentation and technical preservation process. There is a range of issues associated with legal and physical access to digital material. These include:

- the right of individuals to access the cultural heritage of the nation
- the right of creators and copyright holders to protect their commercial interests in the work
- the right of specific cultural and religious groups to restrict access to sensitive material
- the need for organisations to protect both their commercial interests and confidentiality

Finding the balance of rights with the preservation imperative is a continuing challenge. This has serious implications as the protection of commercial interests for creators and publishers in particular not only restricts access but also inhibits preservation processes.

There seems to be a trend in the global online publishing industry in particular towards licensing rights to use online publications and then tightly restricting access to these resources through licensing contracts or other rights management procedures. Libraries are increasingly subscribers to of licensees of digital resources rather than repositories for physical holdings. This has serious implications for ongoing access to all the material that has been licensed from publishers. For example, online publications are licensed rather than sold to libraries such that once the subscription runs out or is no longer current access to past information may be lost. While in an ideal system the material would be archived and properly managed by the licensing company, this brings up the significant issue of long term access.

One of the key issues to be addressed in the management of access to digital material is negotiating a balance between the needs of creators and publishers to safeguard their commercial and legal interests in the material and imperatives for the preservation of digital cultural and intellectual heritage. Pam Gatenby of the National Library of Australia highlighted the need to develop collaborative programs that in combination with legal deposit will assist in addressing preservation needs at the point of creation.
How meaningful access can be provided to a body of material in the future is a critical question. This must be seen in terms of how those with a need to access the material will use the information in the future and what kind of access is going to be needed for this material. This encompasses a range of requirements including rights management, search infrastructure, collections methodologies and data management. To understand these aspects of access to material the problem must be approached at philosophical, academic and technical levels.

SOCIAL (BROAD CONTEXTUAL ISSUES)

The digital agenda is not only precipitating change in organisational structures but in the learning environment and the way that knowledge and information is used in that environment. In The Battle of the Book in the Digital World Clifford Lynch expressed the exciting changes that have occurred in communication:

And completely left behind in the focus on reading technologies, control of intellectual property, and the economics of publishing (and all the broader social implications) is the deep, important, and exciting question of how the digital medium may permit authors and readers to reconceptualize the acts of communication and documentation that have been embodied in the printed book for some or all of the purposes that the book has historically served. This may be the area with the greatest promise of truly transformative changes.

Digital technologies have changed the way we create, store, use and interpret information. Linearity is no longer the most effective method of communication in a text document and dynamic documents are conveying information as printed text could not do. There is necessary link between the way data is physically stored and its interpretation. Geographical barriers are becoming less relevant and globalisation has immense implications for the use of data. Seamus Ross writes ‘Our culture itself is being transformed. The internet has created an environment in which new communities and social groups can evolve, as well as protocols and etiquette governing virtual social interaction’.

The internet is evidence of a major shift in global communication and social structures. It is the memory of its own reforming phenomena. The internet is far more than the system on which it runs and the data which it contains. Since its inception the internet has become an environment of its own, reinventing social experience and the concept of community. Its constantly changing form offers a huge challenge in the preservation of its essence, form and content. The majority of internet sites are dynamic or operate with a database behind the interface. Many sites use interactive mechanisms such as chat as a central feature and it is this material that forms the basis of the social revolution of the internet. It is also a constantly changing environment. Ross notes this saying ‘This virtual world is changing so fast, that behaviours that were evident and observable five years ago, have disappeared because the environment has shifted the behavioural goal posts. The internet and the cultural experiences it provides both reflects real communities and permit us to experience worlds not otherwise open to us’.

The PANDORA Project at the National Library of Australia is the most comprehensive archiving project in Australia. This is an archive of significant online publications copied with the publisher’s permission and preserved and made available for the future. The National Library is one of the leading organisations in preservation of digital material in the world, but there are constraints on the quantity and frequency of material that is archived in PANDORA, set out in its collecting policy.
As noted above, there are countries, such as France, that plan to harvest their entire web domain but this creates its own problems of access, storage and cost. Finding a solution to providing meaningful access to vast quantities of data is one area of concern. Additionally, in the globalised information environment geographical boundaries are becoming redundant. Many collaborative programs require sophisticated access and rights management for the information to continue to be an accurate or authentic record of the original work. If the whole domain is not archived then the process of selecting representative material to archive is also a problem.

Ross addresses the issue of the volume and usability of digital archives, arguing for a maximum collection methodology rather than selection. He approaches the issue of access in a visionary way, saying that ‘Increasingly, research efforts to interpret these data area assisted by data visualisation tools. In addition to data mining and visualisation tools, future researchers will be aided by intelligent agents that explore the Internet (or its successor) looking for information that meets certain user-specific criteria and refining their searches as they accumulate data and knowledge. Digital archives will liberalise scholarship. They will enable simultaneous access to a range of sources (both local and distant) and facilitate the use of research methods not possible with conventionally printed or hand written records’.  

In Avoiding Technical Quicksand: Finding a Viable Technical Foundation for Digital the point is made that what is a useful transformation to a casual reader ‘may be a disastrous loss to a scholar, historian, or lawyer’. Jeff Rothenberg identified the principal concern with the collections management of digital material, writing:

> The essence of preserving information artefacts is the retention of their meaning. This requires the ability to recreate the original form and function of a document when it is accessed, for example, to establish its authenticity, validity, and evidential value and to allow the document's user to understand how its creator and original viewers saw it, what they were (and were not) able to infer from it, what insights it may have conveyed to them, and what aesthetic value it may have had for them.

In particular the key point is the linking of data management with the concept of the original intent of the creator. As noted, the concept of the original in digital material is somewhat hazy. The verifiably original document is crucial for research in a range of government, academic, corporate and private sector functions. Rothenberg writes of the fragility of meaning as ‘the meaning is in the eye of the beholder’ and what may seem a trivial change in digital material for one user may be disastrous for the scholar or historian reliant on access to the authentic object.

The question about how the internet is going to be effectively archived while respecting the authenticity of the material and accessibility for users remains unanswered. There are projects which are planning to archive the internet such as the French government project which plans to take regular snapshots of specific domains; organisations trialing archiving whole domains such as the Swedish Web Archiving Project; and those archiving whole domains such as Alexa in the United States. These projects, however, are still limited to representative rather than comprehensive archiving of all changes and interactions that occur online. It is not feasible for a project, organisation or government agency to take on the task of tracking all changes in a given country’s web domain. Enabling creators and internet site managers to archive their sites is an important step in effective preservation which is being approached by agencies including the National Library and Archives in Australia. A layered system in which archiving occurs at a number levels may create a safety net in archival preservation of the internet but it may not be financially feasible for smaller creators, producers and publishers to archive and preserve the archives of their work.
The maintenance of citations or hyperlinks within documents is another major area in providing ongoing authentic access to online material. Documents may rely on information held in another document that is linked to a specific timeframe. This creates a significant problem in preserving an original online document as its value is often linked to its context set by its citations. Maintaining vigilance in updating links in one way that this is being addressed but this does not guarantee ongoing access. Alternatively, archiving all associated documents may solve this problem but creates issues of rights management to be resolved.

This is being approached by projects such as NEDLIB. This is a collaborative project of European national libraries which aims to construct the basic infrastructure upon which a networked European deposit library can be built to enable ongoing access to online publications. There is an emphasis in such projects on developing communities of information creators, custodians and users. A unique feature of the internet, however, is the change in traditional modes of information exchange. Many creators who make up the diversity of the online environment fall outside the parameters of such communities.

Data management of digital materials in inextricably linked to the use of the data. The management of access mechanisms for the future is vitally important for future scholarship. A key issue here is the way that the data will be used in the future. The use of digital material in the past few decades has revolutionised our understanding of knowledge and information and the way that we use it. By viewing preservation in terms of the more linear tradition associated with the preservation of physical objects are we condemning the future of data to our vision of it? What are the mechanisms by which the data we are preserving is going to be the most valuable and usable it possibly can be? Ross wrote ‘Valuable cultural data contained in record structures, software and hardware will hold keys to understanding the material itself, processes of work, and the culture which created the material’. Essentially the issue is: should we be questioning how the material being preserved is likely to be used in the future to inform our preservation policies, procedures and strategies in the present? In an environment where selection may be necessary for financial, administrative, legal or collection management reasons, selection of material become a critical exercise. The establishment of significance assessment criteria is highly important but given the rapidly changing environment there needs to be considerable work done on developing these criteria in view of the needs of future generations who will access and use it.

**PRESERVATION FRAMEWORKS**

There is an emphasis in preservation programs on the infrastructure framework required to support preservation. This was articulated by the 2001 RLG/OCLC Report that built on the concept of the need for deep infrastructure by defining an ideal scenario of a trusted repository. It set out to address the need for ‘a definition and consensus on the characteristics of a sustainable digital repository for large-scale heterogeneous collections held by research libraries and archives’ and developed a report which describes a framework for reliable repositories. The framework set out for a proposed trusted archive has the following attributes:

- **Administrative responsibility**, a fundamental commitment to implementing the range of community agreed standards and best practices and particularly those that influence viability and sustainability
• **Organisational viability**, commitment to long-term retention and management of and access to digital cultural assets including transparent business practices and management policies

• **Financial sustainability**, a repository with a solid, auditable business plan in place and good business practices and long-term financial planning practices

• **Technological suitability**, including appropriate hardware and software for acquisition, storage, and access, policies and plans for technology management, compliance with standards and best practices, adequate staff expertise, and accessibility for regular external audits of system components and performance.

• **System security**, all systems will be designed to ensure the security of all digital assets managed there including copying processes, authentication systems, firewalls and backup systems, disaster preparedness, staff training.

• **Procedural accountability**, a repository will be accountable for relevant policies and procedures and processes should be transparent and trackable.38

The key concepts are here are access, responsibility, authenticity, accountability and sustainability of management systems for digital materials. The key to achieve this vision of the trusted repository is the framework within which it is set. While this report is somewhat visionary the key elements of a trusted digital repository form the basis of understanding the need for infrastructure to support digital preservation.

In Australia the key proponents of digital preservation are the National Library of Australia (in association with a number of the state libraries), the National Archives of Australia and the National Film and Sound Archive. These organisations have taken different approaches in the preservation of digital material. In particular, the preservation of digital materials has been approached by the National Library (with the state libraries) archiving of online resources considered of high national significance and the National Archives taking an advocacy approach to digital preservation in the government sector.

What does the notion of stakeholder mean to digital preservation? It has implications for education programs, preservation infrastructure (including skills development and resources allocation) and most importantly the concept of responsibility for preservation that is so important with digital material. The stakeholders in digital preservation in Australia are almost all those with a stake in the preservation of the cultural and intellectual heritage of the nation. In particular the stakeholders can be identified as: government bodies, the corporate and academic sectors, software vendors, creators and publishers, libraries and archives, networked information service providers and the digital material audience of present and future generations.

A key theme running though the projects and programs in digital preservation is the need for a culture change in the infrastructure and relationships between organisations at a local, national and international level. Preservation of digital materials cannot be effective if it is not fully integrated into the mechanisms of management at each level of governance. The digital agenda is challenging traditional organisational practices and demanding a fundamental shift in the working practices of organisations adapting to the demands of digital preservation. Existing relationships between departments and between institutions are being reconfigured, roles and responsibilities are being redefined, team work and collaboration at departmental, institutional and international levels is being seen as a solution to a major crisis in records management and there is an increased emphasis on accountability.38 Established systems of archiving and records management are being constantly challenged. There are demands for a skilled workforce that currently does not exist.
In the Australian context digital preservation is a new concept in many organisations. Awareness building, skills development and organisational structures are critically important. This relates to need for firmly established rights and responsibilities to be defined at all levels from government, to the institution, to the individual creator. The most important question here is who has the ultimate responsibility for preservation of digital material, for developing policies, strategies and procedures, implementing best practice standards, collecting and managing the material? Who must decide what is kept, how it is kept, negotiate rights management, locate and manage the resources and develop the procedures and policies that will dictate these activities? Theoretically, responsibility lies with all those involved with the material from the point of creation. Pragmatically, however, preservation practices need to institutionalised for them to be successful in the digital world.

Finally, allocation of resources for the development and sustainability of systems of digital preservation is a major issue of consideration in preservation programs. There are substantial and increasing costs associated with of collecting and maintaining digital material archives and providing ongoing access to the material. The National Library estimated in 1999 that collection of a digital publication alone (without factoring in ongoing maintenance) was five times as labour intensive as for the print equivalent. In 1998 Margaret Hedstrom flagged the need for trusted systems which follow rules and are accompanied by appropriate combinations of policies and standards with the aim of balancing risk, benefit, and cost. This raises the question of how organisations can be structured such that there are adequate available financial and human resources devoted to digital preservation. In a climate where financial resources are scarce and organisations have to prioritise their funding the cost of digital preservation is high.

Alan Howell of the State Library of Victoria raised an important point on the allocation of funding to digital preservation programs (both digitising physical objects and preservation of ‘born digital’ objects). For organisations like the State Library of Victoria it is likely that the limited funds available for the high cost of preservation of digital material require prioritising of available resources. It was suggested that it is not until digital preservation is part of everyday preservation procedures, rather than a special project, that organisations will be able to meet the needs of digital preservation on an appropriate scale. It is the ‘normalisation’ of digital preservation practices that will necessitate the development of policies and procedures, standards and the dedication of human and financial resources for the creation, maintenance and long term preservation of digital objects. The question that must be raised here, however, what is the cost of prioritising of digital preservation over conventional conservation of physical material. The sustainability of digital preservation programs in the long term must also be considered when conceptualising the national, local and institutional infrastructure for digital preservation across sectors.
SUMMARY

In summary, there are a number of key areas for research and development in digital preservation in Australia. These are addressed below by key subject area.

1. Authenticity and the concept of the original

The concepts of authenticity and originality are areas requiring further research and development. In particular, it is the philosophical question of what constitutes the ‘preservable essence’ of a digital resource that still needs to be addressed comprehensively. This could be framed around the following areas of investigation:

- Development of a working definition of authenticity;
- Definition of the significant properties of digital material required to maintain authenticity and assessment of how its significant properties effect the use and access of the material;
- Discussion of the point at which an object is no longer considered original and how much it can be changed by the preservation process before the integrity or authenticity is compromised;
- Resolution of problem of what constitutes the original document in dynamic resources such as internet sites;
- Definition of the meaning of authenticity for different sectors (ie government, corporate, education etc)
- And, review of the technical mechanisms need to be in place to ensure the authenticity of the archived material.

2. Access

Continuing meaningful access to digital materials is the core concern that drives preservation. The need for physical (or technical), legal and intellectual access to authentic digital materials is universally acknowledged as crucial to future users from academia, law, government and politics, among others. Areas that need to be considered research programs in the area of access to digital materials are:

- Negotiating the balance between the need to preserve and provide access to the digital heritage of the nation, and respect for the rights of creators and copyright holders to protect their commercial interests in the work;
- Impact of global online publishing monopolies on access to subscriber material and ways of establishing ongoing access;
- Establishment of a model for a distributed national collection infrastructure that would facilitate maximum access to digital materials within the restraints of creator/publisher rights;
- Development of a strategy for maintaining original citation hyperlink documents so ensure authenticity of online documents. Infrastructure development models may perhaps be along the lines of the NEDLIB project;
- And, in the longer term an issue that needs to be resolved is developing systems for access when whole country domains are harvested.
3. Rights Management

One of the key issues to be addressed in the management of access to digital material is negotiating the needs of creators and publishers to protect and exploit their commercial and legal interests in the material, and the imperatives for the preservation of digital cultural and intellectual heritage. In particular, the development needs in this area include:

- Awareness building programs to encourage creators and publishers to either archive themselves or deposit with the appropriate archiving body
- Development of agreements between stakeholders for deposit of digital materials in the absence of legal deposit, such as the agreement being negotiated by the National Library of Australia and the Australian Publishers Association

4. Collaborative frameworks

There is a need for the development of a sustainable national research and development infrastructure or network to:

- Build collaborative preservation research programs.
- Develop formal collaboration between creators, software vendors, and archiving institutions to raise awareness of the importance of preservation
- Increase education and awareness building programs directed at creators to encourage creators to consider preservation needs at the point of creation
- Develop case studies of working models for trusted repositories to assess feasibility of models such as the trusted digital repository and the OAIS Model
- Undertake a feasibility study on establishing a framework with a layered structure in which archiving occurs at a number of levels creating a safety net in archival preservation

5. Significance Assessment

Currently in Australia archiving is undertaken on a selective basis. While whole country domain web harvesting is being considered, the key concern of access to the vast quantity of data this would create has not be solved. If selection of significant digital material continues to be the chosen method of preservation at a national level then it needs to be complemented by more localised preservation strategies. More work needs to be undertaken on a sectoral basis on establishing significance assessment policies, both generic and specific, with clearly established priories.

6. Data management for future access

The management of access mechanisms for the future is extremely important for future scholarship. Research and analysis may be used as a management tool for future development in data management. Issues that could be addressed include:

- Analysis of possible shifts in the way data will be used and accessed in the future;
- Identification of the mechanisms by which the data we are preserving is going to be the most valuable and usable;
- Questioning how our understanding of the future use of digital material informs our current preservation policies and procedures
- The establishment of significance assessment criteria is highly important but given the rapidly changing environment there needs to be considerable work done on developing this criteria in view of the needs of future generations who will access and use this material.
7. Organisational frameworks
Digital preservation demands a high degree of organisational change. In Australia, there are a number of key issues in enabling organisations to archive their digital materials. Possible areas of research include:

- Development of clearly established chains of responsibility at a national, local and institutional level;
- Establishment of a framework for resource development and prioritisation in records management and preservation programs. This may include education of organisation who could be approached to fund such programs;
- Skills gap audit in digital preservation in Australia and development of strategies to build the skills base in key areas;
- Analysis of awareness of preservation and development of programs to address this in key sectors;
- And, a survey and analysis of the ongoing barriers to preservation at an institutional and sectoral level.

8. Responsibility
Who must decide what is kept, how it is kept, negotiate rights management, locate and manage the resources and develop the procedures and policies that will dictate these activities? Theoretically, responsibility lies with all those involved with the material from the point of creation. Pragmatically, however, preservation practices need to institutionalised for them to be successful in the digital world. An important matter that needs to be addressed is the location of ultimate responsibility for preservation of digital material, for developing policies, strategies and procedures, implementing best practice standards, collecting and managing the material. While responsibility is established for the government sector by archives legislation, it needs to be clearly established for other sectors.

9. Resource development
The allocation of resources for the development and sustainability of systems of digital preservation is a major issue of consideration in preservation programs. Issues that need to be considered are:

- How the substantial and increasing costs associated with collecting, maintaining and providing access to a digital material archive is going to be met in the long term;
- Sustainability of digital preservation programs in the long term;
- How can organisations be structured to ensure that there are adequate available financial and human resources devoted to digital preservation?
- And, how will the normalisation of digital preservation within organisations impact on other organisational activities, including the preservation of physical objects.

10. Priorities in research and development
The research and development agenda, such as is discussed above, must be prioritised in terms of immediate, short and longer term needs. Prioritising is crucial for the allocation of the limited human and financial resources required for such programs.

11. Auditing mechanisms
Finally, criteria need to be established by which the success or otherwise of digital preservation programs can be monitored and assessed.
ACKNOWLEDGMENTS

I would like to thank the following people for generous assistance in the preparation of this paper: Pam Gatenby, Jennifer Hodgement and Gerard Clifton from the National Library of Australia; Adrian Cunningham from the National Archives of Australia; Alan Howell from the State Library of Victoria and Professor Richard Slaughter, Australian Foresight Institute, Swinburne University of Technology and Derek Whitehead, Swinburne University of Technology.

NOTES

3 A selection of these papers are available at www.swinburne.edu.au/lib/DigCon2001.htm
5 Ross, 2000, p.12
6 Ross, 2000, p.12
11 OAIS, 2001, p.iii
12 Hedstrom, 2001, p.8
14 Jones and Beagrie, 2000, p.39
15 Jones and Beagrie, 2000, p.38
19 Pam Gatenby, 20 September 2001
21 Pam Gattenby, 20 September 2001
22 Gatenby, 20 September 2001
23 Jones and Beagrie, 2000, p.16
24 Consultative Committee for Space Data Systems, p.3-1
26 Ross, 2000, p.3
27 Ross, 2000, p.3
28 Ross, 2000, p.22
29 PANDORA, pandora.nla.gov.au/background.html
30 Colin Webb in ‘Towards a Preserved National Collection of Australian Digital Publications’, www.nla.gov.au/nla/staffpaper/webb6.html, December 2000, stated that PANDORA would be highly selective as this was the only way to ensure the quality control over the ‘patchy’ collecting process, negotiate with rights owners, develop a cooperative relationship with creators and publishers, maintain an intimate working knowledge of new web design features and file formats, realistically record the metadata required for future preservation decisions and realistically commit to maintaining access to the diverse and complex files in the archive.
31 Ross, 2000, p.12
33 Rothenberg, 1998, Section 3.
34 See Hedstrom, 2001
35 See Hedstrom, 2001 and NEBLIB www.kb.nl/coop/nedlib/
36 Ross, 2000, p.20
37 RLG-OCLC, 2001, p.3
39 Jones and Beagrie, 2000, p.62
40 This has been addressed by the National Archives in their Guidelines and the National Library through their development of recommendations for creators and publishers with the APA.
42 Hedstrom, 1998, p. 44-71
43 Alan Howell, State Library of Victoria, interview 27 September 2001

BIBLIOGRAPHY


