

# **Digital records preservation in the NSW public sector: a discussion paper**

November 2007

© State of New South Wales through the State Records Authority, 2007. This work may be freely reproduced and distributed for most purposes, however some restrictions apply. See the copyright notice on [www.records.nsw.gov.au](http://www.records.nsw.gov.au) or contact State Records.

## Table of contents

<b>1 Purpose .....</b>	<b>4</b>
<b>2 Introduction.....</b>	<b>4</b>
<b>3 Definitions.....</b>	<b>4</b>
<b>4 The digital records preservation challenge.....</b>	<b>6</b>
4.1 Quantity and variety of digital records .....	6
4.2 Key threats to digital records.....	7
4.3 Digital State archives at risk.....	8
4.4 Government’s goals for better management of information.....	8
4.5 A collaborative effort.....	8
<b>5 Digital records preservation requirements.....</b>	<b>9</b>
5.1 The need for authenticity .....	9
5.2 Legislative requirements affecting digital records preservation.....	11
Evidence Act 1995.....	12
5.3 The size and complexity of the New South Wales public sector.....	13
5.4 Technologies in use now for making and managing digital records.....	14
5.5 The need to preserve both State archives and temporary value records ....	15
State archives.....	15
Temporary value records.....	16
<b>6 Essential tools for digital records preservation .....</b>	<b>16</b>
6.1 Recordkeeping metadata & preservation metadata.....	16
6.2 Model for sustainable repositories .....	17
<b>7 Approaches to digital records preservation .....</b>	<b>18</b>
7.1 Bitstream preservation.....	18
7.2 Encapsulation .....	19
7.3 Emulation.....	19
7.4 Migration.....	20
Format migration.....	20
Normalisation.....	20
Software migration.....	21
Migration on request.....	21
<b>8 Conclusion.....</b>	<b>22</b>
<b>9 References .....</b>	<b>22</b>

## 1 Purpose

The purpose of this discussion paper is to:

- describe the challenges associated with preserving authentic and accessible digital records
- identify the Government of New South Wales' requirements in relation to the preservation of digital records
- explain and evaluate tools and techniques that are used for preserving digital records, and
- propose a next step for the NSW Government to address this issue.

## 2 Introduction

It is vital for the Government of New South Wales to have accessible and authentic records of its business – regardless of their format. Such records are needed for accountability, legal, business and other purposes including documenting our heritage in the State' archives. A record's format should not prevent it from being used or relied upon for these purposes.

As the State's records and archives authority, State Records' aims are to:

- ensure that the business of the NSW public sector is properly documented and that official records are managed efficiently and effectively to support frontline service delivery, good governance and accountability, and
- develop, document, preserve and make available the State archives collection as a cultural and information resource in order to enrich the people and communities of NSW.

Therefore it is our role to lead the NSW Government in committing to the effective preservation of its official records in digital formats; whether they reside in public offices or, in the future, in dedicated digital repositories for digital State archives.

## 3 Definitions

**ADRI:** Australasian Digital Recordkeeping Initiative. This initiative is made up of each of the national, state and territory public records institutions in Australia and New Zealand. The primary objective of ADRI is to pool resources and expertise to find better ways to ensure that digital records are preserved and made accessible for the future. ADRI members have agreed to collaborate on the development, articulation and implementation of a common set of strategies for enabling the making, keeping and using of the digital records of governments.

**Distributed management:** Distributed management is a strategy whereby a public office, or other person, can enter into an agreement with State Records to have possession or custody of State archives. State Records retains control of the archives, while they are held in the custody of the public office or person, ensuring their proper management and care through the

terms of the agreement.

Digital records:	'Born digital' records such as emails, web pages or database records, plus scanned versions of paper records that have been digitised in business processes.
EDRMS:	Electronic document and records management system
InterPARES:	International Research on Permanent Authentic Records in Electronic Systems
JPEG:	a commonly used standard method of compression for photographic images
MPEG:	a video and audio compression standard
ODF:	an open, XML based document file format for office applications that create and edit documents containing text, spreadsheets, charts and graphical elements <sup>1</sup>
Open Archival Information System (OAIS):	an archive, consisting of an organisation of people and systems, that has accepted the responsibility to preserve information and make it available for a designated community <sup>2</sup>
Open source software:	refers to a program in which the source code is available to the general public for use and/or modification from its original design free of charge <sup>3</sup>
PDF:	Portable Document Format. A proprietary but open standard file format for representing two dimensional documents.
PDF/A:	Portable Document Format / Archival. This standard defines a format for the long-term archiving of electronic documents and is based on the PDF standard
Recordkeeping metadata:	data that describes the context, content and structure of records and their management through time <sup>4</sup>
State archive:	A State record that the State Records Authority of New South Wales has control of under the State Records Act 1998. ( <i>State Records Act 1998</i> section 3(1))
State record:	Any record, made and kept, or received and kept, by any person in the course of the exercise of official functions in a public office, or for any purpose of a public office, or for the

---

<sup>1</sup> OASIS ODF Adoption Technical Committee, *Open by Design: The advantages of the OpenDocument Format (ODF) An OASIS White Paper*, December 2006.

<sup>2</sup> Consultative Committee for Space Data Systems, *Reference Model for an Open Archival Information System (OAIS)*, January 2002.

<sup>3</sup> Webopedia, accessed 10 May 2007  
[http://www.webopedia.com/TERM/o/open\\_source.html](http://www.webopedia.com/TERM/o/open_source.html)

<sup>4</sup> AS ISO 15489-2002 *Records Management*, Part 1 Clause 3.12

use of a public office. (*State Records Act 1998* section 3(1))

Temporary value records:	Records with no archival value that are destroyed after approved retention periods.
TIFF:	a file format mainly used for storing images, including photographs.
XENA:	XML Electronic Normalising of Archives. This is the National Archives of Australia's software for converting digital records from their original format into preservation formats. XENA, and the plugin architecture being developed for it are available as open source software. The National Archives has also developed open source software for recording the digital preservation process (Digital Preservation Recorder or DPR) and for providing access to digital archives (Quest).
XML:	Short for eXtensible Markup Language, a specification developed by the W3C. XML is a pared-down version of SGML, designed especially for Web documents. It allows designers to create their own customized tags, enabling the definition, transmission, validation, and interpretation of data between applications and between organisations <sup>5</sup>

## 4 The digital records preservation challenge

The challenge faced by the New South Wales Government of preserving digital records of its business is significant both in scale and complexity. The quantity and variety of records in digital form generated in government today continues to grow, with an ever expanding array of formats in use. These records are vulnerable to risks including loss of authenticity, inadequate description for retrieval and loss of accessibility owing to technological change.

### 4.1 Quantity and variety of digital records

New South Wales Government organisations are increasingly conducting business online and keeping records of that business in digital form.

Digital State records managed by New South Wales public offices come in formats including:

- text files
- web pages, dynamic web content
- digital photographs
- digitised images of hard copy records
- geospatial data
- technical drawings (CAD)
- web content, and
- digital audio / video.

---

<sup>5</sup> Webopedia, accessed 10 May 2007  
<http://www.webopedia.com/TERM/X/XML.html>

Any one of these may need to be retained by the public office that created or received it, as an official record. Many will need to be retained for long periods of time to meet legal and business requirements, and some will be identified as having continuing value and form part of the State archives.

The quantities involved are huge. The results of *State Records' 2005 Information Survey on Digital Recordkeeping*<sup>6</sup> indicated that approximately 630 gigabytes of digital records were being managed in electronic document and records management systems (EDRMS) in the 64 largest Government agencies. The same study indicated that up to 5.5 terabytes of information was being stored in business applications – such as specialised case management systems – in these same 64 agencies. Electronic mail systems generate large quantities of records with agencies surveyed reporting that they receive up to 25 million emails per year, with between 60% and 80% of these relating to their official business. Since 2005 there has been significant growth of digital records as some NSW public offices are creating records, which were previously created as paper records, as strictly born digital. These large quantities of digital records, if not properly managed and preserved in a way that ensures their ongoing authenticity and accessibility, could present a serious liability to the New South Wales Government over time.

## **4.2 Key threats to digital records**

Just as paper based records can be damaged or lost as a result of poor storage, digital records are vulnerable if a proactive approach to their preservation is not taken. Key threats for these records are:

1. a loss of authenticity if they are not adequately protected and controlled,
2. becoming unreadable as software applications and hardware change and
3. possible physical deterioration of the storage media.

Records that are not captured into official systems and managed as records in those systems lose authenticity. For example, without the 'read only' controls that a digital recordkeeping system applies to records; the content of those records could be tampered with. Without the contextual information provided by recordkeeping metadata linked to a record, its proper meaning cannot be understood. A loss of authenticity means that records cannot be relied on as evidence of the organisation's activities, potentially leading to breakdowns in business process, inability to prove matters in the courts, or citizens losing trust in vital government information about their rights and entitlements.

As the New South Wales public sector continues to move towards fully digital business processes, many public offices report that they are not confident that all digital records they create will remain accessible for as long as they are required to be retained. In the *2005 Information Survey*, 38% of respondents indicated that they had some digital records that were either only accessible with some difficulty or no longer accessible at all.

---

<sup>6</sup> The 2005 Information Survey on Digital Recordkeeping was carried out with a small pool of NSW public sector organisations across all sectors covered by the *State Records Act, 1998*. As this survey pool comprised small to very large organisations and was representative of the broad categories within State Records' jurisdiction, findings of the survey were extrapolated to gain broader indications of what might be occurring across the sector. For the full report on the survey go to:

[http://www.records.nsw.gov.au/recordkeeping/information\\_surveys\\_7256.asp](http://www.records.nsw.gov.au/recordkeeping/information_surveys_7256.asp)

Anecdotal evidence in New South Wales shows that public offices are keen to retain digital records instead of the 'print and file' options that many currently employ, but are concerned about these issues of authenticity and accessibility.

By providing public offices with the means to preserve authentic long term digital records, they will be given greater confidence to move to digital recordkeeping and away from storing large quantities of paper based records – a practice which currently represents a significant cost to government.

### **4.3 Digital State archives at risk**

81% of public offices surveyed in the 2005 Information Survey indicated that they wished to transfer at least some of their digital records to State Records' custody as State archives<sup>7</sup>.

A State archive is a record from a New South Wales public office that has been authorised to be retained as part of the State's archival collection. State archives form an essential part of our State's cultural heritage as well as documenting the rights and entitlements of citizens and recording the actions of government to allow the public to eventually scrutinise those actions.

State archives are identified as such because of the activity they document, not their format. So naturally there will be some State archives in digital formats. At present there is no means for State Records to take these archives into custody and preserve them for use by Government now and society into the future. The threat of a 'digital dark age'<sup>8</sup> is a very real concern to those with an interest in the preservation of the memory of the Government of New South Wales.

### **4.4 Government's goals for better management of information**

This challenge is not just State Records' concern. Poor preservation of digital records including those required as digital State archives is a barrier to the NSW Government achieving the goal under its *People First ICT Strategy* of making information easier to access and share with and across government and improving business intelligence and decision support.

### **4.5 A collaborative effort**

Fortunately the challenge of digital records preservation is not one we need to face alone. In 2004 a coalition of government archival authorities from Australia and New Zealand formed the Australasian Digital Recordkeeping Initiative (ADRI) The primary objective of ADRI is to pool resources and expertise to find better ways to ensure that digital records are preserved and made accessible for the future. ADRI members have agreed to collaborate on the development, articulation and implementation of a common set of strategies for enabling the making, keeping and using of the digital records of governments.

Collaboration on this problem is not limited to the archival community. Work done in the library world, using digital repositories for research, such as PANDORA Australia's web archive, established by the National Library of Australia and work within the universities sector can help to further inform State Records' digital records preservation program.

---

<sup>7</sup> Ibid.

<sup>8</sup> Huxley, John 'The Digital Dark Age', *Sydney Morning Herald*, September 23, 2005

## 5 Digital records preservation requirements

The New South Wales Government's approach to digital records preservation must support the keeping of authentic and accessible records for use now and by future generations. It must therefore be shaped by requirements for archival records, legislative rules and by environmental issues specific to the New South Wales public sector. The approach chosen must be:

- feasible – public offices and State Records must have hardware and software capable of implementing the method
- sustainable – either the method can be applied indefinitely into the future or there are credible grounds for asserting that another path will offer a logical sequel to the method, should it cease being sustainable
- practical – implementation should be within reasonable limits of difficulty and expense, and
- appropriate – for the types of objects to be preserved and the specific objectives of preservation.<sup>9</sup>

Accordingly, various New South Wales government requirements influencing our choice of digital records preservation approach are discussed in this section. These are:

- the need for authenticity
- legislative requirements affecting digital records preservation
- the size and complexity of the New South Wales public sector
- technologies in use now for making and managing digital records, and
- the need to preserve both State archives and temporary records.

### 5.1 The need for authenticity

A key requirement for the preservation of digital records in particular – given their transient nature and vulnerability to compromise - is authenticity.

Government and the community expects that the custodian of its archives will protect and preserve them in such a way that they can be relied upon to reflect an accurate picture of Government's activities and can be trusted as 'official', conveying the meaning and intentions of the records creators. This expectation can only be met by ensuring the authenticity of the records being preserved.

Archives and records professionals have been considering how to preserve authentic digital records for some time. A significant project that examined this question was the International Research on Permanent Authentic Records in Electronic Systems (InterPARES), coordinated by the University of British Columbia, Canada. An InterPARES *Preservation Task Force Report* from 2001 notes that:

"Naively, preservation may be seen as a process that keeps records free from change. However, it can be easily shown that it is practically impossible for any record to remain absolutely unaltered or immutable over time.. ..The archival

---

<sup>9</sup> Thibodeau, Kenneth, 'Overview of technological approaches to digital preservation and challenges in coming years', *The State of Digital Preservation: An International Perspective, Conference proceedings* (Washington DC: Council on Library and Information Resources

requirement for integrity, and therefore authenticity, depends on the message intended by the record creator.”<sup>10</sup>

This is a widely accepted view in archival thinking. Digital records preservation programs need to be concerned as much with defining how they will preserve the *intent and meaning* of the record as they are with preserving its *content*. This is no easy task as the characteristics that give a record its meaning can vary from case to case.

For example:

- in a Word document, meaning might be conveyed via bolding of certain words;
- in a digital map, meaning might be conveyed via the colours used;
- in a report from a database meaning might be conveyed in the way the data is laid out.

In their green paper *An Approach to the Preservation of Digital Records* (2002), the National Archives of Australia describe the concept of a record’s ‘essence’ – the combined characteristics of a record which are essential to understanding how its creator meant for it to be understood. The National Archives of Australia’s digital preservation program, now operational, seeks to determine and preserve the essence of the digital records in their custody so they can adequately recreate their meaning for future users. However they do not propose to do this by recreating precisely the original ‘look and feel’ of every record. By contrast, the Public Record Office Victoria’s VERS Project regarded the original appearance of documents to users as a significant consideration in the selection of their preservation technique, in part as a result of the requirements of local evidence laws.<sup>11</sup> The New South Wales *Evidence Act 1995* does not impose such requirements (see section 5.2).

The InterPARES Project’s findings emphasise the importance of the creation and capture of metadata as a fundamental support to a record’s authenticity, including metadata that should come with records, documenting the immediate context of their creation and the manner in which they have been handled and maintained. The findings also indicate that the preserver (such as an archival institution) must be able to demonstrate that:

- an unbroken chain of custody of the records is maintained
- security and control procedures are implemented and monitored
- the content of the record and any required annotations and elements of documentary form remain unchanged after preservation, and
- they have documented the preservation process.<sup>12</sup>

Some of these requirements will be met by enforcing strict controls over the security of the transfer process. Implementing standards such as the *New South Wales Recordkeeping Metadata Standard* and the Australasian Digital

---

<sup>10</sup> InterPARES *Preservation Task Force Report*, Appendix 6: ‘How to preserve authentic electronic records’ [http://www.interpares.org/ip2/ip2\\_products.cfm](http://www.interpares.org/ip2/ip2_products.cfm)

<sup>11</sup> Public Record Office Victoria, *Victorian Electronic Records Strategy Final Report*, 1999. <http://www.prov.vic.gov.au/vers/pdf/final.pdf>

<sup>12</sup> InterPARES *Authenticity Task Force Report*, Appendix 2: ‘Requirements for assessing and maintaining the authenticity of electronic records’, March 2002. [http://www.interpares.org/book/interpares\\_book\\_k\\_app02.pdf](http://www.interpares.org/book/interpares_book_k_app02.pdf)  
[http://www.interpares.org/ip2/ip2\\_products.cfm](http://www.interpares.org/ip2/ip2_products.cfm)

Recordkeeping Initiative's (ADRI) *Digital Record Export Standard*<sup>13</sup> will assist us in specifying that information which we will require at a minimum for preserving authentic digital records.

In summary:

1. Adequate recordkeeping metadata must be linked to or otherwise associated with any digital record being preserved.
2. The approach to digital preservation used must be:
  - capable of demonstrating that the content of the record and essential characteristics remain unchanged as a result of preservation processes
  - capable of recording an appropriate audit trail of preservation activities linked to or otherwise associated all records that are preserved
3. Compliance with digital recordkeeping standards supports authenticity.

## **5.2 Legislative requirements affecting digital records preservation**

There are a number of pieces of legislation that contain provisions which must be considered in the selection of an approach to digital records preservation, including:

- the requirements of the *State Records Act 1998* relating to the making and keeping of records by public offices and the management of the State's archives
- the requirements of the *Evidence Act 1995* for the provision of copies of records from electronic systems, and
- the requirements of the *Electronic Transactions Act 2000* for the retention of records of business conducted electronically.

### **State Records Act 1998**

Under section 14 of the Act, public offices are obliged to 'maintain accessibility to equipment / technology dependent records', for as long as they are required.

The Act also requires public offices to make and keep full and accurate records. The *Standard on Full and Accurate Records*, issued in support of this requirement was developed to reflect the qualities that a record should possess as described by the Australian and International Standard AS ISO 15489-2002 *Records Management*, that is, that records must:

- be made
- be accurate
- be authentic
- have integrity, and
- be useable.

---

<sup>13</sup> The ADRI Digital Record Export Standard prescribes the physical representation of electronic records that are to be transferred to an archive. It specifies the metadata that must be transferred with each file, optional metadata and the way the record is physically represented as an XML document. Available in draft form from [www.adri.gov.au](http://www.adri.gov.au)

These requirements need to apply to records both before and after they have undergone digital preservation processes in any public office. The key requirements in relation to records going through preservation processes are authenticity, useability and integrity. The minimum compliance requirements for these are:

Records must be authentic:

- records are routinely captured into official recordkeeping systems;
- appropriate metadata is created and captured, or otherwise associated with, records.

Records must have integrity:

- unauthorised access, alteration, deletion or destruction of records is forbidden by office policy and practice;
- recordkeeping systems and storage facilities are designed, implemented and monitored to protect records from unauthorised access, alteration, deletion or loss;
- records are uniquely identified;
- migration of records from one system to another is controlled and documented.

Records must be useable:

- records are linked to the business context;
- records relating to the same business activity or transaction are linked to each other;
- location and use of records is recorded and tracked;
- records are accessible for as long as they are required.

In addition, the State Records Act, along with the *Freedom of Information Act 1989* and the *Privacy and Personal Information Protection Act 1998* provide for public access to government records meaning that records must be accessible and useable for as long as they are required.

## **Evidence Act 1995**

This Act, mirrored in most other Australian jurisdictions, abolished the 'best evidence' rule, meaning that copies of records from electronic systems may be provided as evidence in courts of law. The responsible public office must, however, ensure the records' integrity so they will not be challenged by the opposing side. Measures recommended for ensuring digital records will not be regarded as poor evidence includes:

- written policies and procedures should be in place to define operations for development, maintenance and use of electronic recordkeeping systems;
- training and support should be provided to reinforce policies and procedures to staff;
- system controls should be developed to ensure the accuracy and reliability of the records produced;
- system audit trails should be implemented to show the history of the record, i.e. the creation, use and context;
- routine and regular tests of the systems performance should be conducted, including hardware and software;

- adequate security should be provided to limit access.<sup>14</sup>

### **Electronic Transactions Act 2000**

The *Electronic Transactions Act 2000*, section 11 (2) says:

'If, under a law of this jurisdiction, a person is required to retain, for a particular period, a document that is in the form of paper, an article or other material, that requirement is taken to have been met if the person retains, or causes another person to retain, an electronic form of the document throughout that period, where:

(a) having regard to all the relevant circumstances at the time of the generation of the electronic form of the document, the method of generating the electronic form of the document provided a reliable means of assuring the maintenance of the integrity of the information contained in the document, and

(b) at the time of the generation of the electronic form of the document, it was reasonable to expect that the information contained in the electronic form of the document would be readily accessible so as to be useable for subsequent reference, and

(c) if the regulations require that the electronic form of the document be retained on a particular kind of data storage device—that requirement has been met throughout that period.'

These requirements from the Act indicate that where information is made and kept in support of business transactions there is an expectation that its integrity is maintained. There is also a need for that information to be '...readily accessible and useable, both immediately and for subsequent reference'.

In summary:

1. The method of preserving digital records must not damage the integrity, useability, accessibility and authenticity of the records.
2. The preservation process that the digital records go through must be accountable itself.

### **5.3 The size and complexity of the New South Wales public sector**

New South Wales 'public offices' include State government agencies, local government, the public health system, universities and State owned corporations. The New South Wales Parliament, the Courts and the Office of the Governor are covered by the Act by agreement. Over five hundred public offices are included in State Records' compliance monitoring activities.

The sector encompasses a wide variety of organisations of varying sizes. Surveys conducted by State Records have shown that a large proportion of Government agencies and councils are small or very small – small meaning fewer than 80 full time staff, and very small meaning fewer than 20 full time staff<sup>15</sup>. These small and

<sup>14</sup> National Archives of Australia, *Archives Advice 23: Providing records in evidence*, 2003. <http://www.naa.gov.au/recordkeeping/rkpubs/advices/advice23.html>

<sup>15</sup> The survey consulted for this data was the 2004 Records Management Survey available at <http://www.records.nsw.gov.au/recordkeeping/docs/2004%20records%20>

very small agencies are often lacking in technical capabilities and specialised staff in support areas like ICT and records management. While most Government agencies have their head offices in metropolitan Sydney, the majority of councils are located outside of metropolitan Sydney. Of the Universities, more than half are located in regional centres.

Like all other government jurisdictions, New South Wales frequently experiences administrative change, with government bodies merging and splitting and functions transferring from one to the other. This change can often result in dilemmas for those responsible for transferring records from one organisation to another, especially where the records are in proprietary formats.

In summary:

1. The approach taken to digital records preservation must be able to be implemented in public offices of a range of different sizes including small and very small public offices.
2. The approach taken to digital records preservation should not require large investments in additional personnel, technology or infrastructure for public offices choosing to implement it 'in house' to preserve temporary value records.
3. It would be desirable if the approach taken to digital records preservation simplified the process of transferring records between agencies in the event of administrative change.

#### **5.4 Technologies in use now for making and managing digital records**

Despite its size and complexity there is some homogeneity across the New South Wales public sector in terms of applications currently in use. The majority of respondents to State Records' 2005 *Information Survey on Digital Recordkeeping* advised that they used Microsoft Office for their standard desktop applications such as word processing, presentations and spreadsheets. Most agencies use either Outlook (Microsoft) or Groupwise (Novell) for email. SAP is widely used for business process management, financial and HR management, along with specialised products such as CHRIS and Sybase. The New South Wales Government ICT policy *People First* has directed a move towards further centralisation of ICT systems such as websites, exploring standardising 'back office applications' and investigation of open source options for government ICT.

However the New South Wales Government is also characterised by a wide variety of 'home grown' applications built for specific purposes such as case management systems such as the Department of Corrective Services' Offender Information Management System or the Department of Corrective Services' 'KiDS' system.

The nature of systems implemented specifically for keeping records is a critical issue for the success of a digital preservation program. In a survey conducted amongst Government agencies on behalf of the Government Chief Information Office in 2004, it was found that 93% of respondents had implemented a records management system, 47% had implemented a document management system and

---

[management%20survey%20report.pdf](#) This survey did not include area health services.

40% a web content management system<sup>16</sup>. In terms of records / document management systems, the most commonly used product is TRIM Captura (1) and Context (2), with Objective running at third place behind these. It is important to note, however, that there is still a wide range of other products in use in New South Wales Government, many of which are not on the current GSAS RIMS contract, and that the quality of the implementations of these systems is patchy. It would be unrealistic to expect all digital records that are eligible for transfer as archives to have rich and well defined metadata linked to them. In some cases we will need to accept records that have been managed in less than perfect recordkeeping environments, with minimal metadata.

In summary:

1. State Records' digital preservation approach must be able to cope with a wide variety of formats, including non-documentary formats.
2. State Records' digital preservation approach must be flexible enough to cope with records from EDRM applications, especially TRIM and Objective, and from custom built applications such as case management systems.
3. State Records must be prepared to accept records for preservation as archives that have minimal contextual information with them regarding their management and use.

## **5.5 The need to preserve both State archives and temporary value records**

State Records has a responsibility to both preserve digital State archives and also to support and assist public offices in the preservation of temporary value records, some of which may need to be retained for very long periods of time.

### **State archives**

The *State Records Act 1998* permits archives to be physically stored with State Records or with a public office under a distributed management agreement. The approach we take to preserving archives and other records in digital formats must be able to be implemented under either of these models; that is:

- by a public office physically transferring digital records to State Records to be preserved as State archives, or
- by a public office or regional repository retaining digital records as State archives in their custody.

The documentation of context for digital records brought under the control of State Records as digital archives will need to be subject to the same standards as any other archives under the same control. State Records employs the series system for describing and managing archives, which requires documentation of the functional and administrative contexts that records are created in.<sup>17</sup> In addition, digital State archives will need to be processed and accessible quickly, in order to meet public office requirements for retrieval.

---

<sup>16</sup> *Review of Content Management for NSW Government*, conducted by James Robertson, Step 2 Designs, April 2004, p22

<sup>17</sup> For more information see State Records, *Archives in Brief 52 – Introducing Archives Investigator*  
[http://www.records.nsw.gov.au/archives/archives\\_in\\_brief\\_52\\_2106.asp](http://www.records.nsw.gov.au/archives/archives_in_brief_52_2106.asp)

## Temporary value records

Many records that are not required as State archives have very long retention requirements. For example records relating to the environment, health or utilities and infrastructure. Where these records are kept in digital formats, the State Records Act requires that the responsible public office ensures that the records remain accessible for as long as they are required. Currently, public offices can follow the guidelines on ensuring the accessibility of digital records that are offered in the State Records publication *Future Proof*<sup>18</sup>. These suggest a combination of strategies with an emphasis on migration. There is, however, a need for more practical assistance and tools for public offices that need to maintain digital records over long periods of time. By providing such tools and support, there is also an opportunity to reduce costs that could be incurred by many public offices across government potentially implementing their own digital records preservation projects.

In summary:

1. Records may need to be preserved as digital State archives when they are in a public office's custody, a regional repository's custody or State Records' custody.
2. Preserved digital archives must be able to be managed under State Records' existing archival control systems regardless of their physical location.
3. The method of preservation must allow for fast processing of transferred archives, to ensure access requirements are met as soon as practicable.
4. There is a need for digital records preservation tools and advice for public offices that wish to preserve records 'in house'.

## 6 Essential tools for digital records preservation

Over the last ten years the Australian archives and records community has moved from theory to implementation in the preservation of digital records.

The experience of operational digital archives has shown that there are certain tools that we know we must use to successfully preserve and manage digital records.

### 6.1 Recordkeeping metadata & preservation metadata

"Recordkeeping metadata is data that describes the context, content and structure of records and their management through time."<sup>19</sup>

Recordkeeping metadata is vital for the management and useability of digital records throughout their existence, including as archives. Recordkeeping metadata is generated from the moment a record is created or received, and is added or produced each time a record is saved into a recordkeeping system, used, transferred or accessed. It includes a combination of automatically generated, user defined and inherited information.

---

<sup>18</sup> State Records, *Future Proof: Ensuring the accessibility of equipment / technology dependent records*, 2004.  
[http://www.records.nsw.gov.au/recordkeeping/guideline\\_14\\_future\\_proof\\_5\\_051.asp](http://www.records.nsw.gov.au/recordkeeping/guideline_14_future_proof_5_051.asp)

<sup>19</sup> AS ISO 15489 Part 1 *General* Clause 3.12

Recordkeeping metadata is also crucial in supporting the authenticity of digital records as they move through different environments and even change format. The InterPARES project's *Preservation Task Force Report*, Appendix 6: 'How to preserve authentic electronic records'<sup>20</sup> indicates that a record preserver (for example an archival authority) must gather information (metadata) on the context of a record's creation, the manner in which it has been handled and the essential characteristics that give the record meaning if it is to successfully preserve an authentic copy of the record.

Recordkeeping metadata can also be used to ensure that all necessary information about a record's 'content and essential characteristics' is retained and referenced to ensure the preservation process does not compromise these features.

Preservation metadata is information that supports and documents the digital preservation process. For example, it is metadata documenting custody/ownership, preservation processes, technical dependencies and rights management.

However recordkeeping metadata and preservation metadata are interchangeable, meaning that the same metadata can be used for recordkeeping and preservation purposes. Hence there is no need for separate metadata to be maintained for both purposes. There is no need for separate metadata to be made and kept with digital records for recordkeeping and preservation purposes; the same information can serve as both purposes.

Metadata should be standardised; this assists both the organisations creating and receiving the records with the records' management and ensures that an archive that received the records has adequate metadata for their management and use over time. Some examples of metadata standards are:

- *NSW Recordkeeping Metadata Standard (NRKMS)* (preferred for NSW Government Recordkeeping use)
- *PREMIS: PREservation Metadata Implementation Strategies*

## **6.2 Model for sustainable repositories**

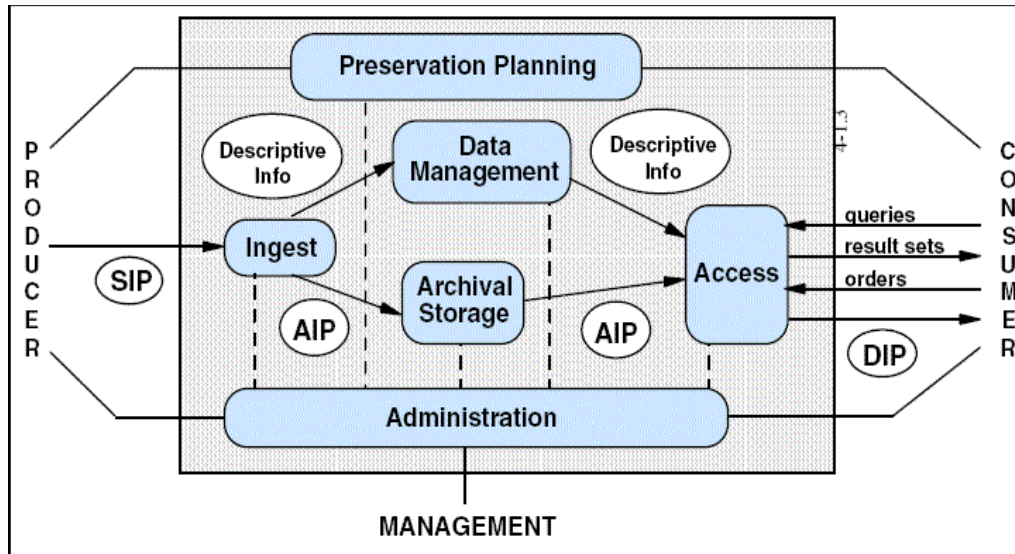
The Open Archival Information System (OAIS) reference model<sup>21</sup> is a proven and widely adopted framework for sustainable digital repositories used by digital archives in Australia and around the world.

### **OAIS Functional entities**

---

<sup>20</sup> InterPARES Project, *Preservation Task Force Report*, 2006  
[http://www.interpares.org/book/interpares\\_book\\_f\\_part3.pdf](http://www.interpares.org/book/interpares_book_f_part3.pdf)

<sup>21</sup> Consultative Committee for Space Data Systems, *Reference Model for an Open Archival Information System (OAIS)*, January 2002



The OAIS model provides a common conceptual framework for describing the environment, functional components and information objects within a system that is responsible for the long term preservation of digital materials. It has been adopted by a number of communities that are concerned with the long term preservation of digital information, including libraries, universities and archives.

Some of the principles underpinning effective preservation of digital materials that the OAIS model conveys are:

- the importance of non technical aspects of a digital repository such as organisational infrastructure and staffing, sustainable resourcing and procedural accountability
- the need for an ongoing program of preservation planning, including routine monitoring of the technological change occurring in the repository's environment, and
- the importance of metadata to record all preservation, management and use processes that take place.

The OAIS model can be implemented in conjunction with different preservation techniques (eg encapsulation, normalisation). Examples of digital repositories in Australia that follow the OAIS model include the National Archives of Australia's and the Public Records Office Victoria's digital archive repositories.

## 7 Approaches to digital records preservation

### 7.1 Bitstream preservation

Bitstream preservation can be used as a foundation for other preservation strategies but is not adequate on its own for ensuring long term accessibility and authenticity. It involves simply storing the binary code (1s and 0s) that comprises a digital object bearing in mind that the object will not be reproducible without the original combination of hardware and software that created it. The advantages of carrying out bitstream preservation include:

Advantages	Disadvantages
Having the opportunity to go back to the	Is not suitable as a preservation

'original' record in this form to carry out different preservation techniques in the future.

strategy on its own.

## 7.2 Encapsulation

In the encapsulation approach, records are packaged as bitstream with metadata enabling a user in the future to display them. The leading example of this approach is the Victorian Electronic Records Strategy (VERS), the digital preservation program of ADRI member the Public Record Office Victoria.

In the VERS approach, record content is accepted in formats including Text files, PDF, PDF-A, JPEG, TIFF and MPEG, encapsulated using an XML 'wrapper' containing a standard set of metadata elements and authenticated using a digital signature. Each record that is 'encapsulated' can contain multiple documents that together form a record.

It is a similar approach to emulation, without the need to include specifications to exactly rebuild the original hardware and software to 'play' the record. Rather, the metadata provides a hardware and software independent method for understanding the record over time. In this sense it is similar to other approaches such as the National Archives of Australia's XENA 'normalisation' technique (discussed below); the difference being the manner in which the metadata is captured, linked to the record and stored.

Advantages	Disadvantages
Content and contextual information kept together to minimise risk of loss.	Can be 'records-centric' - not as effective for recording contextual information about people, organisations and functions.

## 7.3 Emulation

The approach of emulating the original system consists of keeping the original executable program that was used to create and/or manipulate digital information. A specification is also kept that records the details of the original environment that are required to 'play' the program on another computer in the future.

While emulation has been used extensively in other disciplines - for example in gaming, to recreate older computer games for current day players - there are fewer documented projects using it for digital records preservation. There are currently no members of the Australasian Digital Recordkeeping Initiative (ADRI) are currently using emulation as an approach to the preservation of digital records.

Advantages	Disadvantages
Has the potential to be more effective for preservation of databases and multimedia.	Still relatively untested in digital records preservation.

## 7.4 Migration

### Format migration

Using archival data formats is an approach which is usually implemented in conjunction with other approaches such as encapsulation or migration.

A common format used in preserving digital information is XML (eXtensible Markup Language). XML provides a standard syntax for identifying parts of a document known as elements, and then a standard way (known as a schema) for describing the rules for how those elements can be linked together in a document. It is a widely accepted and fully documented way of structuring documents that is supported by many different open source software applications.<sup>22</sup>

A standard that is increasingly being adopted by governments and others to ensure interoperability, ease of access and longevity of digital information is ODF or OpenDocument Format. ODF is "...an open, XML based document file format for office applications that create and edit documents containing text, spreadsheets, charts and graphical elements."<sup>23</sup> ODF is designed to be vendor and implementation neutral, making it possible for people to access, use and share documents regardless of applications or operating systems they are using, so they are not bound by the license they may or may not hold, or the hardware they use. ODF can be used with open source applications such as OpenOffice, which offer the same kinds of desktop applications that are found in Microsoft Office, for example.

### Normalisation

The National Archives of Australia has adopted a particular form of migration called 'normalisation'. This involves migration of digital records to a limited number of standard formats on their arrival at the Archives. At the heart of their approach is the software application Xena (XML Electronic Normalising Archives). Xena detects the file formats of digital objects and converts them into open formats for preservation.

Native formats that XENA can convert include:

- MS-Word, Excel, Powerpoint and Project
- OpenOffice.org Writer, Calc, and Impress
- RTF
- PST email format
- TRIM email format
- MBOX email format
- Comma Separated Files (CSV)
- JPG, GIF, TIFF, PNG, BMP, PCX

---

<sup>22</sup> National Archives of Australia website, Why does the archives use XML? Accessed online May 7 2007, [http://www.naa.gov.au/recordkeeping/preservation/digital/xml\\_data\\_formats.html](http://www.naa.gov.au/recordkeeping/preservation/digital/xml_data_formats.html)

<sup>23</sup> OASIS ODF Adoption Technical Committee, *Open by Design: The advantages of the OpenDocument Format (ODF) An OASIS White Paper*, version last updated 10 December 2006. [http://www.oasis-open.org/committees/download.php/21450/oasis\\_odf\\_advantages\\_10dec2006.pdf](http://www.oasis-open.org/committees/download.php/21450/oasis_odf_advantages_10dec2006.pdf)

- HTML
- Plaintext (various encodings)
- PDF documents, and
- XML.

The National Archives of Australia is conducting ongoing research and development work to expand the list of formats that XENA can recognise and convert.<sup>24</sup>

<b>Advantages</b>	<b>Disadvantages</b>
Data formats which are open standards or which have published codes allow records to be reconstructed if applications are lost.	Converting to a different format may cause the record to lose authenticity if essential characteristics are affected.
XML based formats have popularity worldwide for information sharing and ease of access as well as longevity.	Methods of converting some digital records to archival formats still to be developed.
Tools for converting records to XML formats are now available as open source software	

### **Software migration**

This type of migration is a valid approach for maintaining accessibility and authenticity of records over time while those records are required for current business or while they are being retained for short to medium periods of time.

<b>Advantages</b>	<b>Disadvantages</b>
Migrating records forward as systems change can be made a routine part of a public office's normal ICT upgrades.	Migration poses a risk of loss or alteration of records if not properly managed.
Records are available in current formats with up to date interoperability with other systems.	Can be costly if performed many times over a record's life.
Can be used to maintain records in complex database / case management type systems.	

### **Migration on request**

This approach involves preserving the bitstream of the record and developing a tool which will be capable of reproducing the intellectual content of the record in a different format. The tool must be developed before the record becomes obsolete. Migration is then only performed when a record is requested

<b>Advantages</b>	<b>Disadvantages</b>
-------------------	----------------------

<sup>24</sup> XENA and the plugin architecture being developed for it are available as open source software for download from the website SourceForge (<http://xena.sourceforge.net/index.html>).

Limits the possibility for data loss or alteration from multiple migrations.	Extra effort is required to keep migration tools up to date.
	Risk that the migration tools may themselves become obsolete.

## 8 Conclusion

A policy statement should be developed to formalise the NSW Government's commitment to the preservation of records of Government business in digital formats. This policy should provide a framework for digital preservation that is:

- low in impact and manageable for public offices
- able to support a distributed model for the State's digital archives
- in line with accepted standards for digital preservation such as OAIS, and
- flexible to account for evolving technology and preservation methods.

The policy should address issues including:

- preferred preservation technique/s,
- the responsibilities regarding the defining of essential characteristics and content of digital records during the preservation process, and
- required or recommended practice in managing the preservation process.

The policy should be applicable to any preservation activity carried out on digital State records, whether by public offices or State Records itself.

The policy will serve as an essential foundation for future efforts by State Records in assisting public offices with the preservation of their digital records and in the establishment of the State's first digital archiving facility.

## 9 References

AS ISO 15489 – 2002, *Records Management*

Consultative Committee for Space Data Systems, *Reference Model for an Open Archival Information System (OAIS)*, January 2002.

*Electronic Transactions Act 2000(NSW)*

Ellis, Judith A. (Ed.) *Selected Essays in Electronic Recordkeeping in Australia*, Australian Society of Archivists Inc. O'Connor, Canberra, 2000.

*Evidence Act 1995 (NSW)*

InterPARES Authenticity *Task Force Report*, 2002 Appendix 2: Requirements for Assessing and Maintaining the Authenticity of Electronic Records  
[http://www.interpares.org/book/interpares\\_book\\_d\\_part1.pdf](http://www.interpares.org/book/interpares_book_d_part1.pdf)

InterPARES *Preservation Task Force Report*, 2001.  
[http://www.interpares.org/book/interpares\\_book\\_f\\_part3.pdf](http://www.interpares.org/book/interpares_book_f_part3.pdf)

National Archives of Australia, *An Approach to the Preservation of Digital Records*, 2002.

[http://www.naa.gov.au/recordkeeping/er/digital\\_preservation/Green\\_Paper.pdf](http://www.naa.gov.au/recordkeeping/er/digital_preservation/Green_Paper.pdf)

NSW Department of Commerce, *People First – A new direction for ICT in NSW* July 2006. <http://www.gcio.nsw.gov.au/ict-strategic-plan>

OASIS ODF Adoption Technical Committee, *Open by Design: The advantages of the OpenDocument Format (ODF) An OASIS White Paper*, December 2006

Rothenberg, Jeff *Avoiding Technological Quicksand: Finding a Viable Technical Foundation for Digital Preservation*, January 1998.

<http://www.clir.org/pubs/reports/rothenberg/contents.html>

*State Records Act 1998*

State Records, *Standard on Full and Accurate Records*, 2004.

[http://www.records.nsw.gov.au/recordkeeping/full\\_and\\_accurate\\_records\\_443.asp](http://www.records.nsw.gov.au/recordkeeping/full_and_accurate_records_443.asp)