



Digital Preservation

illuminating the past, guiding the future



Australian Government



NATIONAL
ARCHIVES
OF AUSTRALIA



First published June 2006

© Commonwealth of Australia 2006

ISBN 1920807 45 4 online

This work is copyright. Apart from any use as permitted under the *Copyright Act 1968*, no part may be reproduced by any process without prior written permission from the National Archives of Australia. Requests and inquiries concerning reproduction and rights should be directed to the Publications Manager, National Archives of Australia, PO Box 7425, Canberra Business Centre ACT 2610.

Message from the Director-General

This is a digital age, an era characterised by rapid technological change begun in the second half of the 20th century. Most of us interact with digital technology on a daily basis at work and in our private lives.

As the official custodian of Commonwealth records, the National Archives of Australia has a significant role to play in ensuring the preservation and accessibility of digital records well beyond their creation.

In 2002 the Archives determined that digital records of archival value, created in any format, were to be preserved. This principle has guided our research and development since, culminating in a working prototype of a digital archive.

The cornerstone of our achievement is the software Xena, developed in the open source environment. Xena (XML Electronic Normalising of Archives) converts digital records into a format that can be preserved and used in the future, so they are accessible regardless of technological change.

I hope that the innovative work of the National Archives Digital Preservation team assists others working in this field and that together we can ensure our digital heritage survives into the future.

Ross Gibbs
Director-General
National Archives of Australia



Guiding digital preservation for the future

More than 30 years ago the National Archives of Australia accepted the transfer of about 600,000 computer tapes from the Bureau of Mineral Resources, substantially boosting its fledgling digital collection. The tapes recorded underwater explosions that occurred during petroleum exploration surveys in Australian coastal territory.

By transferring these tapes to the National Archives, the Australian Government intended to secure the long-term preservation and future use of a rich store of geophysical data. What a rarity these computer tapes were in the sea of Australian government paper records routinely dealt with by the Archives in the 1970s!

Digital records are no longer a rarity. Individuals, governments, businesses, and institutions routinely create and capture digital information. Access to personal computers (PCs) and the Internet has fostered expectations of unlimited and perpetual access to records and information. These expectations, and the technical challenges posed by digital records themselves, require the National Archives to explore new ways of preserving and maintaining them.

... Access to personal computers and the Internet has fostered expectations of unlimited and perpetual access to records and information ...



Opportunities and risks in the digital environment

The rapidly changing digital environment has offered many new opportunities. Among these are easier access to records online and the creation of multiple copies of records for back-up and recovery in case of disaster. But with these opportunities comes risk: digital records are more vulnerable than records created in more traditional ways.

Digital records consist of a sequence of encoded numerical values rather than human-readable text. They are commonly stored on physical carriers that will eventually – and sometimes rapidly – become obsolete, like the analogue computer tapes the National Archives received in the 1970s. The carriers are more transient than traditional record formats, such as paper and analogue film. If a paper record fades, it may still be legible. If a part of a digital record stored on an out-of-date medium is corrupted, it is unlikely that the record will be recoverable or readable.

To date the National Archives has been able to preserve records stored on paper and film by preserving the physical carrier

– the medium on which they are recorded. However, this is not sustainable in the digital environment because it requires that both the hardware and software used to create digital records be preserved in order to access the records.

For example, one might want to read a digital document created 20 years ago. Originally the record might have been created on a personal computer running DOS (a common operating system at the time) using something like WordStar, an early word-processing application. The document would have been saved to a 5¼” floppy disk that required a 5¼” disk drive, which is not a feature of contemporary PCs. Once removed from the system in which it was created, this digital record is ‘orphaned’ – it becomes very difficult to open and read.

This example highlights the main obstacle facing the National Archives when preserving digital information: the rate at which information technology is developing has resulted in rapid obsolescence of equipment and software.



Obsolescence: a key challenge in the digital age

Digital records are subject to three types of obsolescence:

1. ***The physical carrier of the record becomes obsolete.***
Standard media of the 1980s, such as 8" and 5¼" floppy disks, are no longer commercially available. Current media such as CDs and DVDs will also eventually become obsolete.
2. ***The hardware needed to access the record becomes obsolete.*** Both the drives needed to read the media and the computers required to operate them have become obsolete in our rapidly changing world. It is almost impossible to obtain working 8" and 5¼" floppy disk drives. Most new computers have a usable life of only three to five years and are not fitted with floppy disk drives.
3. ***The software needed to access the record becomes obsolete.*** This includes both the software needed to read and write the record and the operating system on which that software ran. For example, during the 1980s WordStar was widely used, but it is now no longer readily available. The fact that WordStar used a closed data format makes continuing accessibility even more problematic.



National Archives preserves records of archival significance

The National Archives selects, maintains and preserves Australian government records of enduring value, and makes them available for public access 30 years after their creation.

The breadth of this responsibility in a rapidly changing digital environment is a real challenge. Hardware and software obsolescence requires the National Archives to actively intervene to maintain digital records and provide access to them long after their creation.

Consider the type and volume of records government agencies create in their day-to-day operations. The Department of Finance and Administration, to select just one example, generates numerous records each day – some are paper but many are digital. They comprise many formats – word-processing documents, spreadsheets, emails, images – dealing with the Australian Government Budget, financial initiatives and reviews, and financial advice and training for all of government.

Some 18 Australian Government departments, with responsibility for 187 agencies, generate a vast number of records each year. Generally, no more than 5 per cent of records are deemed to be of archival significance and transferred to the National Archives for safekeeping and future access. That 5 per cent of records will eventually comprise more digital records than paper records.

The need to operate as a digital archive presents resource, infrastructure and systems challenges for the National Archives. At the same time, there is an ongoing need to maintain a parallel traditional paper archive service.



National Archives' digital preservation approach

The National Archives has developed an innovative approach to meet the challenge of digital preservation. It is now operating a prototype that has been recognised nationally and internationally as a sustainable, scalable and innovative answer to the complexities of digital preservation and access.

The National Archives' approach is to convert digital records into open preservation file formats to guarantee access to their contents in the future. The open formats are based on standards, have full specifications that are publicly documented, and are interoperable with a range of software applications.

The Archives has avoided using proprietary (ie closed format) software because the owners of such software tend to keep some or all of their specifications secret and their software is needed to access all aspects of a format. They may also hold patents and may ask for royalty payments where their format is implemented.

The Archives has developed software tools to convert data into open formats, as well as tools to export to original formats and access converted data in the way it was originally presented. The use of open file formats will allow others to build tools capable of presenting or repurposing data preserved by the National Archives.

... The National Archives' approach is to convert digital records into open preservation file formats to guarantee access to their contents in the future ...



Digital preservation process

The National Archives' digital preservation process is founded on the fundamental principle that good recordkeeping and archival systems provide access to complete, reliable and authentic records into the future. The records must be safe from unauthorised access, alteration and deletion. In order to achieve this, the Archives takes into account the need for data security, data integrity and audit requirements. Owing to the increasingly larger quantities of digital records to be transferred to the Archives, the process needs to operate on an industrial scale.

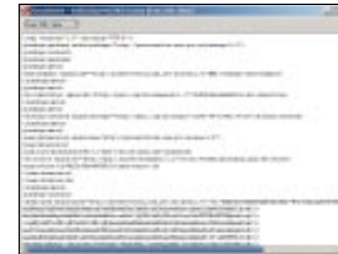
The Archives' digital preservation process has been separated into three distinct phases:

1. **Quarantine.** Records are checked for viruses and completeness against lists provided by government agencies.
2. **Preservation.** Records are preserved (normalised, or converted, into open formats) using the in-house developed software application, Xena (XML Electronic Normalising of Archives).
3. **Storage.** Records are deposited for long-term storage in a digital archive.

Each phase operates in a physically isolated network that is customised to its particular needs. Records travel through each isolated phase of the process until they are eventually stored in the digital archive.



Left. Image presented by the Xena Viewer after conversion from jpeg

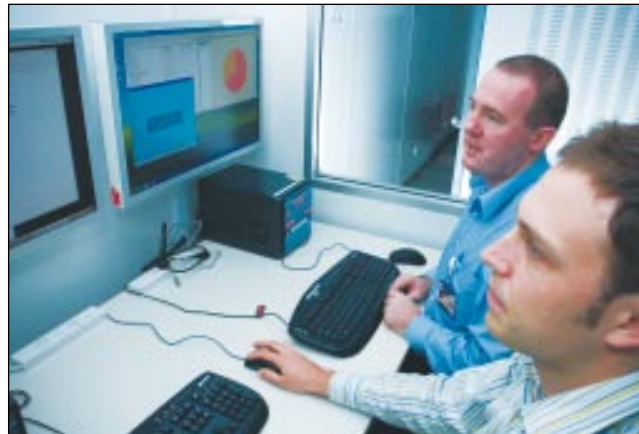


Right. The same image presented as XML by the Xena Viewer

Physical environment and hardware

The National Archives' digital preservation facilities comprise a secure and stable environment that houses processing networks, a digital archive and a separate laboratory for staff. The facilities are located in two access-controlled and network-isolated rooms, with a back-up power supply, environmental monitoring and a fire suppression system. The environmental stability and security features help to ensure the integrity and life of the hardware and software and – more importantly – the records. All digital preservation processing occurs within these two isolated rooms, one for the hardware and software, and the other for staff.

When selecting the hardware and systems for its digital preservation prototype, the National Archives avoided relying on any single vendor or solution. In doing this, the Archives enhanced its ability to deal with hardware obsolescence – technology is the enabler, rather than the driver, of the Archives' processes.



Selecting the right software

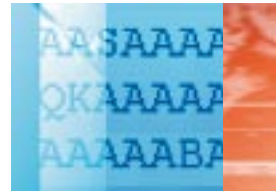
When researching and developing its approach to digital preservation, the National Archives made several key decisions about the software it required:

- The software would be developed under an open source licence (the GNU General Public License or GPL) for transparency, and to encourage its dissemination to government agencies and other organisations needing to access digital records over a long period. In the spirit of open source development, the Archives would establish a network of collaborators – a community of users – who would contribute to the development of the software.
- The processing software would be able to be implemented across platforms, so the programming language Java was chosen.
- The software would convert proprietary file formats into XML (Extensible Markup Language), a widely accepted and fully documented way of structuring documents supported by many different applications. Records in open formats have a potentially greater lifespan in the digital world.

The digital preservation software the National Archives developed is designed to control workflow and manage audit, conversion, storage, retrieval and access.

- **Workflow control and audit.** This is carried out by Digital Preservation Recorder, an application that works across all processing networks to ensure data integrity, authenticity and reliability by recording preservation metadata about the processed records.
- **Conversion.** The conversion of records into XML-based preservation formats is performed by Xena, an in-house developed software application that also enables preserved records to be displayed for access.
- **Storage and retrieval.** Storage and retrieval of records in the digital archive are undertaken with Quest (Query Electronic Storage), an application that enables the management of, and ultimate access to, records stored in the archive.

This combination of process, infrastructure and software enables the National Archives to preserve, for future access and use, selected digital records created by the Australian Government.



Preparing for the future

To keep abreast of changes in information technology and the dynamic nature of digital recordkeeping, the National Archives must undertake ongoing research and development. It will focus on the following areas:

- **Process.** The digital preservation process must evolve in line with changes in the Archives' own business processes and developments in digital records creation, management and preservation.
- **Infrastructure.** Regardless of what the Archives does to minimise changes required in infrastructure, the prototype itself will, over time, become obsolete and require change. In part this will be driven by the Archives' need to scale its operations in line with the storage capacity required to maintain digital records created by government agencies.
- **Software.** The preservation software must not only evolve to meet changes in the process and infrastructure but also keep up with changes in the digital recordkeeping environment. Currently, the Archives converts (normalises) office documents, emails, images and some other files into open file formats, but there are many more digital formats, and more will evolve

... In the spirit of open source development, the National Archives is establishing a network of collaborators ...

in the future, that will have to be normalised and preserved. Xena's plug-in architecture enables the software to be readily enhanced to meet this challenge.

While its digital preservation approach is already in operation as a prototype, the National Archives continues to plan and prepare for a future when the volume of records will require a more sophisticated system to handle significantly larger operational requirements.

As keepers of digital records of enduring significance, the National Archives will continue to promote a comprehensive understanding of how to preserve and maintain such records. It will also refine processes to handle large volumes of these records.



Preservation is a shared responsibility

The National Archives shares responsibility for preserving Australia's digital heritage with the government agencies that create the records. Many agencies already have unreadable legacy digital records. By raising awareness about obsolescence, the National Archives is seeking to ensure that the number of unreadable digital records does not grow.

As for the legacy digital records that already exist, the Archives can support and guide agencies in determining whether the records can be retrieved and if required, preserved.

Both the National Archives and Government agencies must respond proactively and creatively to the challenge of preserving and maintaining Australia's digital records heritage. We have a shared responsibility for guiding digital preservation into the future.

Further information

Publications on the Archives' website www.naa.gov.au

Simon Davis, *Looking Back to the Future: 30 Years of Keeping Electronic Records in the National Archives of Australia*, National Archives of Australia, Canberra, 2004.

Helen Heslop, Simon Davis, and Andrew Wilson, *An Approach to the Preservation of Digital Records*, National Archives of Australia, Canberra, 2002.

Contact

Digital Preservation
National Archives of Australia
PO Box 7425
Canberra Business Centre ACT 2610

Phone: +61 2 6212 3423
Email: recordkeeping@naa.gov.au
Website: www.naa.gov.au



Obtaining Xena

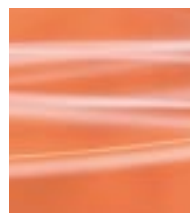
Xena 3.0 Lite is an easy to use and practical software tool that can be used to convert digital data such as office documents into open file formats best suited to long-term preservation. It has been built using the power of the National Archives' Xena Digital Preservation framework and showcases only a selection of the features of the Xena framework. The Xena framework has been designed to easily integrate with other software such as collection management applications and therefore can be adapted to the digital preservation and access needs of other organisations.

The latest version of the Xena 3.0 Lite software may be downloaded from <http://xena.sourceforge.net/download.html>

For more information about Xena, how you can utilise it in your environment and participate in Xena's development, email the National Archives of Australia Digital Preservation team: recordkeeping@naa.gov.au.



+n/7w
AAAAA
BAAS
QOKA
AAAA



Ca/7wMzAl
AAAAAA
AAAAAA
AAAAAA
AAAAAA
AAAAAA
AAAAAA

AAAAQAAAA
AAEAAAAeD
AAAAACOE
AABAFoAAA

