Recordkeeping for Good Governance Toolkit

GUIDELINE 18: Digital Preservation

PARBICA

Pacific Regional Branch International Council on Archives

The original version of this guideline was prepared by the Pacific Regional Branch of the International Council on Archives (PARBICA) for use by countries around the Pacific. This means that the guideline may refer to things that you are not familiar with or do not use in your country. You may find that you need to change some of the advice in this guideline to suit your own government's arrangements. To obtain an editable copy of this guideline, contact the national archives, public record office or other records authority in your country, or contact PARBICA at http://www.parbica.org.

Recordkeeping for Good Governance Toolkit

Guideline 18: Digital Preservation

CONTENTS

WHO IS THIS GUIDELINE FOR?

This guideline is primarily for archives and library staff in the Pacific Islands, whose core business is preserving long-term access to information. The guideline aims to provide these staff with practical advice on the issues associated with digital preservation, and how to address them. A secondary audience is the staff of government ministries, especially records and information management staff and information technology staff. Members of the secondary audience should make sure to read Appendix C of this guideline, How to avoid information loss in the digital age, which is designed as a separate handout for staff of government ministries.

This guideline is structured so that it moves from a discussion of why digital preservation is an important issue, through an overview of approaches to and strategies for digital preservation, to more technical content about file formats and digital preservation tools. The guideline finishes with a glossary, a further reading list, two technical appendixes and an appendix that summarises the key messages for managers and staff of government departments about how to avoid information loss in the digital age. The final appendix, Appendix D, while not about preserving born-digital records, discusses issues and approaches in digitally reformatting analogue audiovisual materials for preservation purposes.

THE FIVE GOLDEN RULES OF DIGITAL PRESERVATION

1. Manage your digital records like your paper records: use structured classification schemes, good folder titling, and good naming rules.

2. Only retain digital records that you need to keep (in other words, know what you need to keep and delete everything else as soon as it is no longer needed).

3. When undertaking preservation actions, only work on one copy.

4. Always back things up.

5. Make friends with your IT colleagues.

PRESERVING DIGITAL ACCESS

Like paper records, digital records need to be preserved to make sure that they can be found, accessed and used for as long as they are needed.

Digital preservation is the range of activities that are carried out to preserve digital information. Digital information may include a range of born-digital records (those originally created in digital format) such as emails, web pages, documents and spreadsheets, as well as reborn-digital records that have been produced from analogue material as part of a scanning or conversion/reformatting project. Digital preservation does not refer to the process of copying analogue material to digital form (known as digitisation), but rather the actions undertaken on the digital files created as a result of digitisation.

The complex and fragile nature of digital records means that there is a significant risk that the data they hold will be inaccessible, even in just a few years' time. On average, digital data have a life cycle of only seven to 10 years.

To view a paper record, all you need is the document (or have a copy) and an understanding of the language it is written in. Digital records are written in computer language, which is incomprehensible to people. In other words, digital records are dependent on technology. Also complicating things, the data in the file are written in different ways (called file formats), depending on the program that creates them. The exact details of the file format are often only known to the manufacturer/vendor of the program that creates it. To access the data stored in a digital record, you require a software program that can interpret the file format and represent it on a computer screen in a way people can understand.

Preservation of digital objects needs to be more proactive than paper preservation. While you can put a book on a shelf and return to it 100 years later and still open it and read it, using the same approach for a digital object almost guarantees that it will not be accessible in the future.

Technological obsolescence

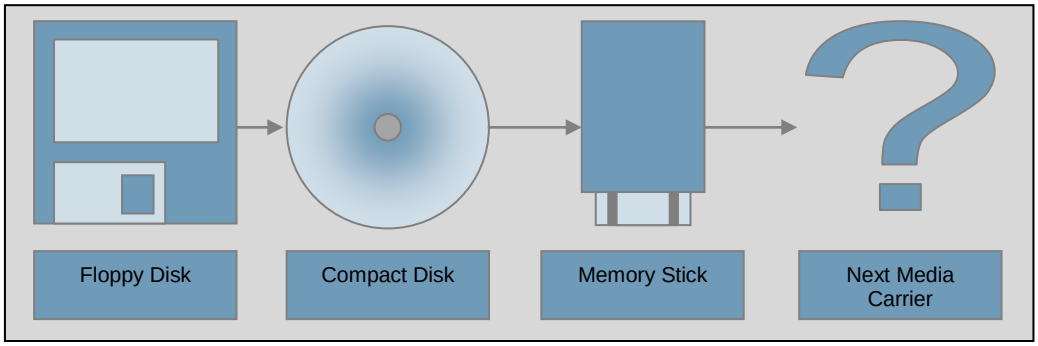
One of the key challenges of preserving digital records is dealing with the fact that computer-based technology goes out of date. This is known as technological obsolescence. When a particular software program is superseded or a new version of it appears, or when a hardware device is no longer produced by the industry, the records created with those technologies may no longer be accessible. Preservation strategies are needed to ensure that digital records live beyond the life of the system on which they are created. To ensure the continuity of digital information it is vital for that information to be well managed from the moment of its creation.

An example of technological obsolescence

You created a document on an old work computer years ago, and saved it onto a 5 ¼-inch floppy disk drive. You now want to view this document but your workplace no longer has computers with floppy disk drives.

You contact a friend who does have the right drive on their computer. You insert a disk and can see your document but you cannot open the file, as the software application you used to create the file is no longer used, and there is no other program that understands the file format.

What could have been done: A solution to the hardware obsolescence problem may have been to continually migrate the file to current media carriers (such as a CD), before your workplace got rid of the computers with floppy disk drives.



To prevent the software obsolescence problem you could have migrated the content from one digital file format to another as each is updated.

PRESERVING DIGITAL AUTHENTICITY

A key digital preservation challenge is ensuring the authenticity of the records over time. Because records are a source of evidence of decisions and activities, it is vital that records can be trusted to be what they claim to be, and that users can trust that the records have not been tampered with, corrupted or otherwise changed. When records are trustworthy and reliable as evidence, and can sustain examination in a court of law, they are authentic. The ease with which digital records can be altered, either deliberately or accidentally, makes authenticity a serious issue. Whereas a paper record contains certain physical characteristics that are obvious indicators of authenticity (such as a letterhead or a signature), these are not always so obvious or fixed on a digital record.

When a record is moved from one system to another (such as during a software upgrade), or from one medium to another (if it is copied from a server to a CD, for example) it is at risk of a loss of integrity and authenticity due to corruption of the files. Refreshing data, implementing good IT practices for data security and back-ups, and maintaining multiple copies of the original bit streams are some practices to attempt to ensure the authenticity of digital objects.

STRATEGIES FOR DIGITAL PRESERVATION

As for paper preservation strategies, there is no 'one-size-fits-all' strategy for digital preservation. A range of strategies that are appropriate to different categories of digital material may need to be used.

Common approaches to digital preservation are:

1. Technology preservation – this approach involves preserving the original software that was used to create and access the information, and preserving both the original operating system and hardware on which to run it. This strategy is not considered feasible over the long term, though it can be useful to keep one or two generations of earlier software and/or hardware.

2. Emulation – this approach involves emulating older software and hardware, using modern computers. Emulation requires significant resources and technical expertise and is generally not a viable option in small to medium-sized organisations.

3. Migration – this approach involves transforming the digital information into new formats before the old format becomes obsolete, while preserving the intellectual content of the information and retaining the ability of users to retrieve, display and use the information.

Of these three approaches, migration is the most reliable, cost-effective and commonly used digital preservation strategy, and is explained in more detail on pages nine to 10 that follow.

Backing up digital information

It is very important to back up your organisation's digital information to prevent the loss of this information in the case of a disaster. Backing up involves copying information onto different storage media, such as a separate hard drive or tapes. *Always hold at least two copies of information, a working copy and a back-up copy, preferably off site.*

Back-ups are very important for making sure that business can continue in the event of an emergency, but they cannot be relied on as a recordkeeping system or long-term archiving strategy. For instance, back-ups do not generally include metadata, and the data may be saved in a single mass which makes location and retrieval of specific files very difficult. There is also the risk that backed-up data may not be accessible over time without the correct software. Where possible, export metadata from digital repositories and EDRMS systems and include these in the back-up. Portable hard drives are very cheap these days, and implementing a simple off site back-up system is relatively easy.

An organisation’s ICT area is generally responsible for backing up digital information. Records managers should speak with their ICT area to make sure that records and documents on back-up media fulfil their purpose and are kept only as long as they are needed.

Managing storage media

The physical media used to store data (such as computer tape, hard disks, etc) are much more vulnerable to physical damage than books and other paper objects. While paper is prone to deterioration or attack by mould or insects, the deterioration is slow and may not become apparent for some decades. For digital objects, however, damage can be quick and instant. One small scratch on a CD can make all the files on it immediately inaccessible; a portable hard drive or laptop can slip from your hands and be damaged beyond repair in a second.

As the quantity of digital documents being created and distributed increases, so too does the range and variety of portable storage media used to store and share digital information. USB sticks, CDs and DVDs are just some examples of media that can be used to store information for back-up or mobile purposes. Portable storage media should not be used to store master copies of records, and should be kept somewhere secure when not being used.

MIGRATING RECORDS

Migration shoud be:

1. planned, with risks assessed and managed.

2. done sooner rather than later, as leaving it may be more expensive or even make migration impossible.

3. done carefully.

4. monitored and documented.

5. checked to ensure that the information has retained its meaning, usability and integrity.

Migration is the transfer of data or a digital resource to a newer operating environment or to an environment that is less likely to be adversely affected by technological obsolescence. This may include conversion from one file format to another (for example conversion of Microsoft Word to OpenDocument Format), from one operating system to another (for example, Windows to Linux) or from one programming language to another (for example, C to Java) so that the data remains fully accessible and functional.

Migration to new operating environments often means that the copy is not exactly the same as the original piece of information. Some functionality, such as macros inside an Excel spreadsheet, can be lost when migrating between file formats (see Appendix A for more information on file formats). When migration results in the loss of information or functionality, it is referred to as ‘lossy migration’.Lossy migration is often more of an issue when the original file is in a proprietary format because tools may not be able to adequately interpret the data structure. The risks related to reading data in proprietary file formats only get worse over time, so it is usually better to migrate as soon as possible to help mitigate this risk.

Before undertaking a migration it is important to assess how much information will be lost during the process and conduct a risk analysis to determine whether the consequences of that information loss are insignificant or significant. Where the information loss is identified as significant pursuing another strategy may be justified. Either way, these deliberations need to be documented for future reference and for auditing purposes, particularly to enable the organisation to justify the authenticity of the information in relation to preservation treatments that have been used.

Migration may be done in a variety of ways. At its simplest, migration may involve copying digital information to a more stable non-digital medium, such

as paper or microfilm. While paper and microfilm may be more reliable in the long term, frequently they don't completely resemble, or have the same functionality as, the original digital object. Transferring files to a newer version of the storage media, or different storage media (for example from floppy disk to CD or to a portable hard drive – sometimes called 'media refreshing') – offers a short- to medium-term strategy for preserving access, but still requires the digital files to be migrated when the technology changes.

Another approach to preserving access to digital information is to migrate it initially to standard formats, which may be more stable than the original formats. The selection of a format for preserving digital information will depend upon what aspect of the resource will be required in the future. For example, a need to process or edit a digital resource in the future or preserve visual presentation will impact upon the format used to preserve digital information.

The complexity of the migration process will vary according to the digital resource being migrated. The migration of an interactive multimedia object will be much more complex than migrating simple text. At its most complex, migration can be time consuming and costly, though it will still be much cheaper than trying to read obsolete data in the future. In any case, migration should be planned and budgeted for as a routine part of ongoing systems and data maintenance.

It is important to remember that like digital records management in general, digital preservation will have ongoing costs. Costs include labour, maintenance of hardware and software, replacement of media and capital equipment, any software licences, storage costs and power costs.

8-STEP DIGITAL PRESERVATION APPROACH

When an organisation commits to ensuring the long-term preservation of its digital information, the following 8-step approach is a guide to implementing a digital preservation strategy.

Step	Activity	Examples
1	Plan	• Link preservation activities with records appraisal to make sure you are not wasting effort preserving records of short-term value. • Build preservation activities into normal records/archives program planning. • Conduct regular monitoring of digital records. • Identify and test the digital preservation tools you want to use.
2	Build partnerships and communicate with your partners	With: • ICT staff, for their technical expertise. • Management for resource allocation. • External experts, archival authorities and digital preservation industry groups.
3	Keep records in systems that can ensure the integrity, authenticity and accessibility of the records	• Records must be controlled and identified to allow of monitoring of accessibility. • Records need to be fixed such that they cannot be altered or tampered with.
4	Use good recordkeeping metadata	Keep and link to records, information on: • the business context in which the records were created. • their technical dependencies, any migration or conversion activities. • essential characteristics that preserve and ensure authenticity, eg a map may need to be preserved in colour.
5	Use open formats	• As far as practicable, use open data formats to make migration easier. • See Appendix A for more information.
6	Manage storage media	• Carry out routine checks of storage media. • Regularly replace media to limit the risk of damage or loss of records. • Store media in a stable, control environment.

7	Migrate – move records through new formats, media and systems	• Keep records on up-to-date storage media, eg a portable hard drive, instead of a 5 ¼-inch floppy. • Bring records forward as versions of software change. • Plan, manage and document each migration.
8	Protect digital assets	• Perform regular back-ups for disaster recovery purposes. • Ensure records are protected from viruses and 'hackers'. • Use password protection and other access controls.

GLOSSARY

Authenticity – a record that is authentic can be proven to

- be what it says it is;
- have been created or sent by the person supposed to have created or sent it; and
- have been created or sent at the time stated.

Check-sum – a number derived algorithmically from the content of a digital document, intended to check the accuracy of transmission, copying or recording of the information content of a digital document.

Digital file – a logical assembly of machine-readable binary data stored within a computer system.

Digital signature – information which, using cryptographic techniques, provides guarantees of the authenticity and/or reliability and/or authorship of a digital file.

Emulation – using programs that imitate the original (obsolete or unavailable) hardware and software used to create a digital file.

File format – the particular way that information is encoded for storage in a computer file and for use by a software application.

Lossy migration – occurs when migration of computer files results in the loss of information or functionality in those files.

Metadata – data describing the context, content and structure of records which enables their discovery, use, management and preservation through time.

Migration – changing the format of digital information so that it can be viewed and used with different, usually newer or longer-term, hardware and software, while maintaining the authenticity, integrity, reliability and usability of the information.

Open source software – computer software that is distributed free of charge under a licensing arrangement and which allows the computer code to be shared, viewed and modified by other users and organisations. Open source software development is often performed by a distributed community of software developers via the internet.

Proprietary format – computer software applications and/or file formats that are developed, owned and controlled by a private commercial entity, where the software code or specification is not readily available and usually cannot be used without paying a licence fee.

Refresh – to copy digital information from one storage media device to another to protect the information from loss caused by deterioration of the original storage media.

Storage media – the physical carrier for digital information, examples include hard disks, digital tapes, portable hard drives, flash drives, compact disks, floppy disks.

Virus – a computer program or code that is transferred to a computer system or digital file without user knowledge and with the intention of corrupting or deleting information in the recipient computer.

FURTHER READING

Brown, Adrian, *Archiving Websites: A practical guide for information management professionals*, Facet Publishing, London, 2006.

Digital Preservation Coalition, *Preservation Management of Digital Materials: The Handbook*, York, 2008, http://www.dpconline.org/advice/digital-preservation-handbook.html

Fleischhauer, Carl, 'Format considerations in audiovisual preservation reformatting: Snapshots from the Federal Agencies Digitization Guidelines Initiative', *Information Standards Quarterly,* 22 (2), Spring 2010, pp. 34 – 40, http://www.digitizationguidelines.gov/audio-visual/documents/IP_Fleischhauer_AudioVisual_Reformatting_isqv22no2.pdf

Harvey, Ross, *Digital Curation: A how-to-do-it manual*, Facet Publishing, London, 2010.

Harvey, Ross, *Preserving Digital Materials*, K.G. Saur, Munich, 2005.

International Association of Sound and Audio Visual Archives, *Guidelines on the Production and Preservation of Digital Audio Objects*, 2nd ed., 2009, http://www.iasa-web.org/tc04/audio-preservation

Jasper, Mike, 'How to convert reel to reel audio to digital', http://www.ehow.com/how_5796206_convert-reel-reel-audio-digital.html

Lawrence, Alexis, 'How to convert old VHS tapes to DVDs', http://www.ehow.com/how_6743273_convert-old-vhs-tapes-dvds.html

Müller, Heiko, *Database Archiving*, Digital Curation Centre, Edinburgh, 2009, http://www.era.lib.ed.ac.uk/bitstream/1842/3346/3/Mueller%20Database%20archiving%20_%20Briefing%20Paper.doc

RLG-NARA Task Force on Digital Repository Certification, *Trustworthy Repositories Audit and Certification: Criteria and Checklist*, Center for Research Libraries, Chicago, 2007, http://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf

APPENDIX A – FILE FORMATS – OPEN AND PROPRIETARY

A file format is the particular way that information is encoded for storage in a computer file and use by a software application. File formats are indicated by an extension that appears at the end of the title of the document or digital object, for example '.html', '.pdf' and '.doc'. The decision of which file formats to use for creating born-digital records should be made with long-term sustainability in mind – as well as any immediate business requirements. It is useful to identify a minimal set of formats that meet both the active business needs and the sustainability criteria, and only create data in these formats.

There are two main types of file formats, *open formats* and *proprietary formats*. To use a proprietary format it is necessary to agree to (and often pay for) licensing conditions imposed by the private company that owns the source code for that format. Open formats are not owned by private owners, but instead are community owned and developed, though to use open formats you still need to agree to abide by licensing conditions. Open formats are also called free file formats if they are not bound by restrictions such as copyrights, patents and trademarks so that anyone may use them free of charge for any desired purpose, subject to certain conditions of use. Although open formats are generally preferable for digital preservation purposes, just because a format is open it does not necessarily mean that it is good. A good indicator of the quality of an open format is the extent of adoption of that format by organisations and members of the community.

Open formats and proprietary formats are explained further below.

	Open format	**Open proprietary format**	**Closed proprietary format**
Examples	Portable Network Graphics (PNG), OpenDocument Format (ODF), JPEG, Extensible Hypertext Markup Language (XHTML), Tagged Image File Format (TIFF), Free Lossless Audio Codec (FLAC), Portable Document Format (PDF) - ISO 32000.	MP3 audio, Microsoft Office Open XML – eg Docx (OoXML, also issued as ECMA standard 376 and ISO 29500 transitional).	Microsoft Word (DOC), Microsoft Outlook, Excel (XLS), PowerPoint (PPT), Photoshop (PSD), Microsoft Access, RAW image formats, WAV audio format.
What is it?	Publicly shared intellectual property, usually maintained by a standards organisation.	Privately-owned intellectual property.	Privately-owned intellectual property.

Availability of the format specification	Specification published without any restrictions.	Specification may be made available with restrictions.	No specification publicly accessible.
How is it developed?	The format is developed through a publicly visible, community-driven process.	The format is developed and marketed by companies which control the way the technology is used, to improve their market position.	The format is developed and marketed by companies which control the way the technology is used, to improve their market position.
How can it be used?	Can be used and changed by anyone without restrictions, except for licensing conditions that may limit development of commercialised versions of software.	Licence holder has exclusive control of the technology to the (current or future) exclusion of others.	License holder has exclusive control of the technology to the (current or future) exclusion of others.
What software is needed to use it?	Open formats are free to be implemented by anyone, including both proprietary and free and open source software, using the typical licenses used by each.	Generally only licensed applications are free to use these formats.	Proprietary file formats can only be accessed using the software that produced that file, or licensed applications.
Conclusion	Open formats, which are supported by a wide range of software or are platform-independent, are recommended for use where possible. Open formats support long-term sustainability of data by allowing migration from one technical environment to another, without locking into a specific vendor.	Open proprietary formats are a greater risk than open formats because they are controlled by a corporate entity under licensing arrangements that may change.	Proprietary formats carry greater risk to long-term accessibility of the data they hold. The lack of documentation of the specification, and licensing requirements for software means the format is less future-proof.

APPENDIX B – TOOLS FOR DIGITAL PRESERVATION

There are many solutions available for harvesting, identifying and managing data – including a number of free and open source tools. All solutions require – at the absolute minimum – one computer workstation, a staff member with computer experience to implement the software and import and manage data, and storage for data (two hard drives at the very least).

Open-source digital preservation tools		
Name	**What does it do?**	**Who might use it?**
HOPPLA (Home and Office Painless Persistent Long-Term Archiving) HOPPLA is free and can be downloaded for free at http://www.ifs.tuwien.ac.at/dp/hoppla/release/index.html#anchor3. Released publicly September 2010	HOPPLA provides a set of different strategies for preservation of digital data, including back-up and migration. HOPPLA copies files to different locations and converts some of these copies to other formats if there is a risk that the original source file will become obsolete. The tool can handle documents, personal photos and/or videos and digital music collections.	Private users and small organisations that know they need a digital preservation system but don't have the skills or knowledge to develop one.
National Archives of Australia's Digital Preservation Software Platform (DPSP) The source code and support documentation for DPSP is free and available at http://dpsp.sourceforge.net/.	DPSP migrates data from its original format into open, fully documented formats for archival preservation. The DPSP tools comprise Manifest Maker (which creates a list of digital files and directs it to a specified location), XENA (which converts files to open format standards), Digital Preservation Recorder (the workflow tool which ingests and stores files), and Checksum Checker (which monitors the contents of the digital archive for corruption). DPSP is a repository system, which interfaces with the National Archives' collection management database, and uses the unique barcode applied to the file by the organisation.	Larger, well-resourced organisations with access to professional ICT skills and good ICT infrastructure.

Archivematica http://archivematica.org/wiki/index.php?title=Software)	Archivematica is a different system to the repository-based DPSP. Instead of using a complex database or external control system that tracks the links between the object and its metadata, it uses one system where the object and metadata stay in the same file system.	The simplicity of Archivematica makes it ideal for organisations with limited financial and technical capacity.
Platform or framework tools (software applications that work as building blocks for a digital archive)		
Fedora Commons (Flexible Extensible Digital Object and Repository Architecture)	Fedora provides a digital asset management (DAM) architecture, upon which many types of institutional repositories, digital archives, and digital library systems can be built. It is widely used, open-source software that can be run on different operating systems, but requires a lot of technical expertise to implement adequately.	Larger, well-resourced organisations with access to professional ICT skills and good ICT infrastructure.
DSpace	DSpace provides the tools for the management of digital assets and a platform for preservation activities. It is also open-source, offers a repository system that can handle a wide range of digital objects, can be run on different operating systems, and has a wide user base – particularly in educational institutions, but less commonly in government agencies.	Educational institutions needing to manage large quantities of digital resources.
Web archiving tools – the most common web archiving technique uses 'web crawlers' – a computer program that browses the web to automate the process of collecting web pages.		
Heritrix web crawler (developed by the Internet Archive, a major digital library)	Supports key processes such as permissions, job scheduling, harvesting, quality review, and the collection of descriptive metadata.	

Web Curator Tool (WCT)	An open-source workflow management application for selective web archiving. It is designed for use in libraries and other collecting organisations, and supports collection by non-technical users while still allowing complete control of the web harvesting process. It uses the Heritrix web crawler.	
HT Track	Allows you to download a world wide web site from the internet to a local directory, arranging the original site's relative link structure. The user can open a page of the 'mirrored' website in their browser, and browse the site from link to link, as if viewing it online.	
Open-source migration tools		
Image Magick www.imagemagick .org	A tool for image migration.	
SoX www.sox.sourgeforge .net	A tool for sound migration.	

TRAC

Organisations creating new digital repositories and wanting to assess the soundness and sustainability of their repositories can use TRAC (Trusted Repositories Audit and Certification: Criteria and Checklist).[1] In general, TRAC:

- Provides tools for the audit, assessment, and potential certification of digital repositories
- Establishes documentation requirements required for audit
- Sets out a process for certification.

[1] Online Computer Library Centre (OCLC) and Center for Research Libraries (CRL), *Trusted Repositories Audit and Certification: Criteria and Checklist*, 2007, http://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf

APPENDIX C

TEN STEPS FOR KEEPING DIGITAL RECORDS USEABLE[2]

This appendix may be used as a handout for managers and ICT staff in government ministries.

Most organisations are creating and storing rapidly growing volumes of digital information – that is, information created using computers. Often this information is vitally important as a source of information and evidence. Loss of this information would seriously impair the ability of the organisation to function efficiently and effectively, and could cause critical damage to the reputation of the organisation among its clients, citizens and stakeholders.

Digital information is highly vulnerable to deliberate or accidental loss or destruction. Rapid changes in technology can render older forms of digital information unusable and inaccessible. It is vital that organisations put in place strategies to ensure that critical information is available and usable for as long as it is needed.

1. Know your records

Know:

- what computerised systems you have and the technologies that they rely on to function, as sooner or later these technologies will become obsolete, making the information stored in the system at risk of loss.
- Know what information is created and kept in each system, the formats in which the information is stored, where it is, how it is described, and why it exists.
- Know how the information relates to the business of your organisation, how it is used and how long it needs to be kept to meet business and legal requirements.

[2] Adapted with permission from State Records Authority of New South Wales advice, 'How to avoid information loss in the digital age', http://futureproof.records.nsw.gov.au/wp-content/uploads/2010/05/Managing-digital-records-leaflet-Final1.pdf, and Queensland State Archives Public Records Brief, 'Keeping records useable – Ten steps for ensuring the continued accessibility of digital records', http://www.archives.qld.gov.au/publications/ PublicRecordsBriefs/ Ten_steps_digital_records.pdf

2. Design systems to support your records

Many computerised systems are not designed to maintain long-term stable access to information. It is important to recognise long-term information use as a functional requirement when designing or purchasing a new system. In cases where the data will need to be kept beyond the expected life span of the system (usually five to 10 years), it is critical that the system is able to export the data in a usable form so that it can be carried forward into future business systems.

3. Limit the number of file formats you use

File formats are the mechanism by which different types of digital information are encoded and stored for use. Because file formats become obsolete over time, it will be necessary to migrate critical business data in old file formats into new or more stable file formats. The more file formats your organisation uses the more expensive and labour intensive it will be to ensure the longevity of your information assets, and the greater the risk of losing vital information will be.

4. Use open formats

Open formats are not owned by any software companies and so are less vulnerable to loss through technological obsolescence resulting from proprietary dependencies. As a general rule open formats are more stable and are easier to migrate. Wherever practicable, use open formats such as PDF, HTML, XHTML, ODF, JPEG and FLAC.

5. Don’t keep digital information any longer than you need to

Keeping digital information indefinitely is expensive and often technically complex. Although digital storage may appear cheap, storage is not the only cost associated with keeping digital information – there are also significant costs associated with managing and migrating the information. Even with cheap storage, the rapidly expanding volume of digital information means that keeping data for longer than necessary will involve unsustainable costs. You need to know how long information needs to be kept and how to find it, and have systems and processes in place that can ensure that the important information is properly preserved while the unimportant information is disposed of in a timely and managed way.

6. Know where all your information is and keep it under control

Organisations may store their information in centralised databases, on portable storage media, on back-up tapes, in personal or shared folders, or it may be managed by contracted service providers. Increasingly commonly, it may be stored in the so-called internet ‘cloud’, which may mean that the information is actually stored in another country. The more diverse and unmanaged your information storage arrangements, the more your information is at risk of loss.

7. Describe your information well

Digital information needs to be well described (using metadata) so that it can be found, used and managed. Good metadata includes: meaningful and accurate titles; information on access and use conditions and restrictions; information that links the data to its business context; and information about how long the data needs to be retained.

8. Reduce duplication

Digital information can be easily copied. Often organisations will have thousands of copies of the same piece of information stored in different places. Managing and storing many copies of the same data is a waste of money and can lead to confusion by staff and clients. Systems and processes need to be put in place that identify and eliminate duplicates, control versions and guarantee the integrity and authenticity of official records.

9. Manage migration

Migrating data from old, soon-to-be obsolete, file formats and storage media to new or more stable formats is a necessary preservation activity for information that needs to be kept for longer than a few years. Migration is a high-risk process that can threaten the integrity and even the existence of important information. Migration projects must therefore be carefully planned and executed to mitigate risks and to protect the authenticity, integrity and accessibility of the information.

10. Don’t leave it until it is too late to ensure the survival of your information – a stitch in time saves nine!

Ensuring the longevity of critical business information is your responsibility today and cannot be left to your future colleagues to retrospectively fix the problems that you have left for them. Always think about how today’s systems and processes can help ensure that critical business information will be available in the future when it is needed. Proper management of your information from the start will save a lot of money and trouble in the future.

APPENDIX D

DIGITAL REFORMATTING OF ANALOGUE AUDIOVISUAL RECORDINGS

Just as digital information can become inaccessible as a result of technological obsolescence, analogue audio recordings, films and videotapes can too. Reel-to-reel and cassette audio tapes, VHS and Beta videotapes and 16mm and 35mm film are all becoming obsolete. Not only are the physical tape carriers deteriorating, industry is no longer actively supporting the formats. It is becoming almost impossible to buy or repair analogue audio and video tapes and the equipment needed to play those tapes. Any organisation that holds valuable analogue audiovisual material will need to transfer the content to a digital format so that it will be accessible into the future.

The International Association of Sound and Audiovisual Archives (IASA) has produced guidelines on the creation and preservation of digital copies from analogue originals (see reading list and see also paper by Carl Fleischhauer). The IASA guidelines cover:

- selection and preparation of best available copy for digitising
- optimising signal extraction
- analog-to-digital converter technical specifications
- target audio format: linear PCM Broadcast Wave Format
- storage recommendations.

The equipment required to digitally reformat analogue audiovisual recordings to archival standards can be very expensive to purchase and operate. Unless your organisation has a very large quantity of such material it is may be advisable to outsource the work of digitally reformatting analogue audiovisual collections. In the Pacific Islands this may mean outsourcing the work to an offshore service provider. As with any outsourcing arrangements, special care and attention needs to be given to performing due diligence checks, negotiating contracts, contract management and quality assurance of the services provided.

Alternatively, relatively cheap software and equipment can be purchased to do this in-house, though the results may be of uncertain and uneven quality. See reading list and articles by Mike Jasper and Alexis Lawrence.

Once audiovisual recordings have been converted into digital form, the resulting digital copies will of course need to be preserved using digital preservation strategies outlined in this guideline.

The Recordkeeping for Good Governance Toolkit is produced by the Pacific Regional Branch of the International Council on Archives with assistance from the National Archives of Australia and AusAID.