

5. Media and Formats

5. Outline

Intended primary audience

Operational managers and staff in repositories, publishers and other data creators, third party service providers.

Assumed level of knowledge of digital preservation

Novice to Intermediate.

Purpose

To outline the range of options available when creating digital materials and some of the major implications of selection. To point to more detailed sources of advice and guidance. To indicate areas where it is necessary to maintain an active technology watch.

5.1 Media

It is important to have an understanding of the various media for storage because they require different software and hardware equipment for access, and have different storage conditions and preservation requirements. They also have varying suitability according to the storage capacity required, and preservation or access needed.

Although it is very easy to focus on the traditional conservation of the physical artefact, it is important to recognise that most electronic media will be threatened by obsolescence of the hardware and software to access them. This often occurs long before deterioration of media (which have been subject to appropriate storage and handling) becomes a problem. However, appropriate selection, storage and handling of media is still essential to any preservation strategy (see **Storage and Preservation**).

Obsolescence of previous storage media has occurred in rapid succession. In floppy disks alone we have seen a progression from 8 in to 5.25 in and then 3.5 in formats, with each change leading to rapid discontinuation of previous formats and difficulty in obtaining or maintaining access devices for them.

Mass storage devices have a long history and this section deals only with the magnetic and optical storage media which are in widespread or recent use. An interesting historical account of "new media" can be found in the PRO Preservation Guide series (**Farley 1999**).

Magnetic media

Consist of a variety of magnetic media and containers including a range of magnetic tapes (e.g. reels, cartridges and cassettes) and disks (e.g. hard disks, floppy disks). They all utilise the magnetic properties of metallic materials suspended in a non-magnetic mixture on a substrate or backing material.

This provides a versatile and cheap storage medium and both the storage capacity and the ability to retain the magnetic charges holding the data have increased substantially in recent years. The method of construction and storing the data also point to potential weaknesses of magnetic media.

You should ensure appropriate storage away from strong magnetic fields as these may alter the media and lead to data loss (e.g. electrical equipment and motors). Damage from magnetic fields is rare and the media normally has to be in very close proximity (<50 mm) for this to occur. Tape enclosures or packing with a space clearance of 50 mm around the media is recommended for use during transportation and transfer.

Clean operating conditions and environments will reduce the scope for damage to media and devices. The high density of storage and the close proximity of device heads to the media mean even small particles such as smoke or other debris can lead to data loss.

Handling and use of magnetic storage media should be minimised to reduce wear, or refreshment cycles implemented (as recommended by the manufacturer) to replace media on a more frequent basis reflecting the levels of use.

Poor environmental storage may also lead to oxidation of the ferromagnetic material or problems with the "binding" layer or substrate materials. Recommendations for the storage environment of magnetic media are provided in **Storage and Preservation**.

Magnetic media are constantly evolving and in addition to fundamental changes in devices manufacturers often undertake an almost constant evolution of production processes. Although the reliability of magnetic media has improved over recent years it is important to be aware that faults in manufacture can occur and to make appropriate checks of new media when purchased. Media should also be of high quality and purchased from reputable brands and suppliers. As an additional safeguard archive copies can be made to comparable magnetic media purchased from different suppliers to guard against faults introduced into products or batches of the product by the manufacturers.

In addition to the magnetic media themselves it is important that attention is paid to the recording and access devices such as tape drives. These should be of good quality and well-maintained. Problems with the access devices e.g. head/media crashes are one of the most common cases of damage to magnetic storage media. It may also be desirable to write archive copies from different devices and software to protect data from malfunctioning devices or software.

Optical media

Optical storage media such as CD-ROM (Compact Disc - Read Only Memory), CD-R (Compact Disc -Recordable), and DVD-ROM (Digital Versatile Disc - Read Only Memory) use laser light to read from a data layer. In CD-ROM this data layer consists of a series of pits and plateaux in a metallic coating over a plastic disk. A clear acrylic coating is applied to the metallic layer to protect it from scratches and corrosion. CD-R employs a dye layer which is light sensitive as the data layer. Data is written to and read back using laser light. The use of

light sensitive dyes means CD-Rs are less stable than CD-ROMs and more concerns have been raised over their use as archival media (**Ross and Gow 1999**). As with magnetic media there is considerable diversity in practice and production of CD-R and greater care is needed in selecting high quality media from reputable suppliers for archival purposes. DVD-ROM is a more recent optical storage medium with capacity to store 4.7-18 Gb.

Optical disks are an increasingly popular method of storage. The device reader is not in contact with the disk and mechanical failure is less likely to lead to data loss than damage to the disk itself through poor handling or storage. Disks should not be flexed or their surfaces marked or abraded e.g. through use of a sharp pen or pencil for labelling. The manufacturer's recommendation for marking should be followed.

As with magnetic media, optical media have been subject to a constant process of evolution and changes in manufacture. The quality of the media, a reputable source, and appropriate handling and storage environment (see **Storage and Preservation**) will all affect its longevity.

Media life

Media should be refreshed on a regular cycle within the lifetime for archival storage identified by the manufacturer or independent sources such as the US National Media Laboratory. Sample generic figures for lifetimes of media under various temperature and humidity levels assuming optimal use (no or very infrequent access) and environmental conditions (stable and free of contaminants, u-v light and strong magnetic fields) are given in the figure below. It should be noted that the life of specific media will be dependent on the quality of manufacture. Media life will vary between specific products and dates (e.g. the earliest CDs will be more experimental in manufacture than current versions; branded "Gold" CDs will have longer life than cheaper standard products).

Figure 7

Sample Generic Figures for Lifetimes of Media

Device	25RH 10°C	30RH 15°C	40RH 20°C	50RH 25°C	50RH 28°C
D3 magnetic tape	50 years	25 years	15 years	3 years	1 year
DLT magnetic tape cartridge	75 years	40 years	15 years	3 years	1 year
CD/DVD	75 years	40 years	20 years	10 years	2 years
CD-ROM	30 years	15 years	3 years	9 months	3 months

After **Dollar**, **NML** and **PRO**

5.2 File Format and Standards

As with storage media there is a diverse range of formats (e.g. Word, TIFF) in common use. The purpose of this section is not to provide a detailed or exhaustive list of current formats for different media types but to draw attention to the broader implications of file formats for their application, and implications for preservation. There are a number of excellent sources of more detailed advice on file formats and these are detailed in the further reading to the chapter.

File formats are subject to similar rapid obsolescence and evolution and the process of selection and assessment of options for preservation is largely one of risk reduction. Use of file formats which have been well documented, have undergone thorough testing and are non-proprietary and usable on different hardware and software platforms minimises the frequency of migration and reduces the risk and costs in their preservation. Similarly utilising formats which have been widely adopted minimises risk as it is more likely that migration paths will be provided by the manufacturers and a degree of "backward compatibility" will be available between versions of the file format as it evolves. It is important to note that backward compatibility is rarely maintained for more than one or two previous versions and that the "window of opportunity" to migrate is therefore relatively brief.

Although such non-proprietary formats can be selected for many resource types this is not universally the case. For many new areas and applications, e.g. Geographical Information Systems or Virtual Reality only proprietary formats are available. In such cases a crucial factor will be the export formats supported to allow data to be moved out of (or into) these proprietary environments.

It is advisable for institutions where possible to identify file formats which are preferred for archival storage and to seek deposits in that form wherever a choice of formats exist. Some institutions have also identified and distinguished between preferred, acceptable and unacceptable formats for transfer to the institution, for archival storage once in the institution's care, and formats which can be provided for users. Narrowing the range of file formats handled streamlines the management process and reduces preservation costs. It will also reduce the ongoing cost of software licences required by the institution (see also **Acquisition and Appraisal** and **Storage and Maintenance**). In considering storage and preservation it is helpful to recognise that it can be a desirable strategy to distinguish between formats (or versions) used for archiving and access on the basis of different requirements e.g. it would be appropriate to store a high resolution image as a TIFF master file (archival format), but to distribute the image as a JPEG file (access format) of smaller size for transmission over a network. It would not be appropriate to store the JPEG image as both the access and archival format because of the irretrievable data loss this would involve.

The speed with which many file formats evolve and the degree to which even well documented standard formats can be extended by proprietary additions or modified/adapted

155

The Handbook was first compiled by Neil Beagrie and Maggie Jones and is now maintained and updated by the DPC.

for specific applications by users also has significant implications for preservation, and in particular for good preservation metadata and system documentation (see **Metadata and Documentation**).

5.3 Compression and Encryption

File compression algorithms can substantially reduce file sizes and have been widely used in document or image transmission. Compression can either be lossless or lossy (with data loss but often higher levels of compression). Although appropriate in many cases for access and user copies, compression adds additional complexity to the preservation process and is normally not recommended for the storage of archival files. With current increases in storage capacity and reducing costs it is also less necessary. For some very large files e.g. digitised video, compressed formats may be the only viable option however for capture, storage and transmission.

In a similar way encryption is increasingly prevalent either to ensure that sensitive data is read only by the recipient or to ensure a digital product can only be used by an authorised user. Encryption also adds to the complexity of the preservation process and should be avoided if possible for archival copies. This may require strict implementation of physical and system security procedures for the archive of unencrypted files, or archival access to encryption keys.

5.4 Technology Watch

An implication of the rapid evolution of storage media and file formats and the risks of technology obsolescence is the necessity of maintaining a register of hardware and software capacity in the institution and preservation metadata to enable a formal process of "technology watch". The degree to which this will be necessary will vary according to the degree of uniformity or control over formats and media that can be exercised by the institution. Those with little control over media and formats received and a high degree of diversity in their holdings will find this function essential. For most other institutions the IS strategy should seek to develop corporate standards so that everybody uses the same software and versions and is migrated to new versions as the products develop. Deborah Woodyard (**Woodyard 1999**) describes how preservation metadata was gathered by the National Library of Australia to determine what hardware and software were required by its digital holdings. A list of hardware and software available in the NLA was also developed and maintained. This is used to flag potential changes in technology and the requirement to retain hardware and software still needed by the collection until migration has occurred.

Failure to implement an effective technology watch or IS strategy incorporating this will risk potential loss of access to digital holdings and higher costs. It may be possible to re-establish access through a process of "digital archaeology" (see **Preservation Strategies**) but this is likely to be expensive compared to pre-emptive strategies.

A retrospective survey of digital holdings and a risk assessment and action plan may be a necessary first step for many institutions, prior to implementing a technology watch.

Good preservation metadata in a computerised catalogue identifying the storage medium (3.5 in floppy disk, DVD etc.), the necessary hardware (IBM PC compatible, Apple Mac), operating system (Windows 95, NT, Dos 3.0 etc.) and software (e.g. Word 6) will enable a technology watch strategy.

5.5 Summary Recommendations

Media

- Keep store and access areas free of smoke, dust, dirt and other contaminants.
- Store magnetic media away from strong magnetic fields.
- Transport magnetic media in enclosures with space clearances of 50 mm.
- Store in a cool, dry, stable and secure environment (see Storage and Preservation).
- Acclimatise media before use.
- Use high quality media and devices.
- Keep access devices well maintained and clean.
- Do not place labels on optical disks and/or mark using a pen or pencil.
- Follow manufacturers' recommendations for labelling.
- Minimise handling and use of archival media and/or record number of accesses/use and implement appropriate refreshing.
- Write archival copies from different devices and software.
- Make archive copies to comparable media purchased from different suppliers.

File formats

- Use "open" non-proprietary, well-documented file formats wherever possible.
- Alternatively utilise file formats which are well-developed, have been widely adopted and are de facto standards in the marketplace.
- Identify formats acceptable for the purposes of transfer, storage and distribution to users (these may be distinct).
- Minimise the number of file formats to be managed as far as is feasible/desirable.
- Do not use encryption or compression for archival files if possible.

Technology watch

- Undertake a retrospective survey of digital holdings, a risk assessment and action plan.
- Implement a process of technology watch and/or implement procedures for standardisation and changes in technology in your IS strategy.

157

The Handbook was first compiled by Neil Beagrie and Maggie Jones and is now maintained and updated by the DPC.

- Maintain a list of hardware/software available within the institution and use this to flag implications for technology change and hardware/software replacement/retention.
- Ensure you have good preservation metadata in a computerised catalogue which can form the basis for technology watch and monitoring.
- Consider "digital archaeology" to retrieve access to data in obsolete formats.

See Exemplars and Further Reading

1. Beagrie, N. and Greenstein, D. (1998). Managing Digital Collections: AHDS Policies, Standards and Practices. Consultation Draft. December 1999.
<http://www.ahds.ac.uk/about/reports-and-policies/index.htm>
Section 2.9.2 Technical Standards, provides a summary of preferred formats recommended by AHDS service providers. Further details are available in individual Guides to Good Practice.
2. DLM Forum (1997). Guidelines on Best Practice for Using Electronic Information.
<http://europa.eu.int/ISPO/dlm/documents/gdlines.pdf>
Update 19 March 2008
No longer available - information at
<http://ec.europa.eu/archives/ISPO/dlm/>

Chapter 5, Short- and long-term preservation of electronic information, offers advice on data storage media (including advice on storage conditions) and file formats. The latter general advice is "Best practice is to decide on a common set of standards from the outset to make it easier to circulate information.

Preferably the same formats should be used for both short-term and long-term preservation". Both storage media and file formats are grouped into families, with examples of the major types in each.

3. Frey, F. (2000). File Formats for Digital Masters.
<http://www.rlg.ac.uk/visguides/visguide5.html>
Update 09 Aug 2006
New location <http://www.rlg.org/legacy/visguides/visguide5.html>

One of five guides commissioned by DLF and CLIR and published with RLG. This guide provides steps in how to select file formats for digital masters, selecting those based on a combination of performance and durability.

4. National Library of Australia. (1999). First Steps in Preserving Digital Publications.
<http://www.nla.gov.au/pres/epupam.html>

5. PADI. This is highly recommended as providing comprehensive links to relevant resources. Relevant sections available online include:
magnetic media at: <http://www.nla.gov.au/padi/topics/59.html>
optical disks at: <http://www.nla.gov.au/padi/topics/53.html>
physical formats at: <http://www.nla.gov.au/padi/topics/52.html>
6. TASI. Framework webpages are highly recommended as a resource and are available online at:
<http://www.tasi.ac.uk/advice/advice.html>

Includes general advice on selecting file formats for images.

7. van Bogart, John (1995). Magnetic Tape Storage and Handling. Council on Library and Information Resources. (ISBN 1-887334-40-8).
<http://www.clir.org/pubs/reports/pub54>
8. Dale, R. (1999). File compression Strategies Discussion at ALA. RLG DigiNews February 15 1999.
<http://www.rlg.ac.uk/preserv/diginews/diginews3-1.html>
Update 09 Aug 2006
New location <http://www.rlg.org/preserv/diginews/diginews3-1.html>

Search Other Resources

Search Archives of Digital Preservation Jiscmail list

<http://www.jiscmail.ac.uk/cgi-bin/wa.exe?S1=digital-preservation>

Search Preserving Access to Digital Information (PADI) Gateway

<http://www.nla.gov.au/padi/search.html>

References

1. Farley, J. (1999). An Introduction to Archival Materials; new media (PRO Preservation Guide series). Available free from the PRO.
2. Ross, S. and Gow, A. (1999). Digital Archaeology: Rescuing Neglected and Damaged Data Resources. British Library Research and Innovation Report 108. London, The British Library, 1999.
<http://www.hatii.arts.gla.ac.uk/research/BrLibrary/rosgowrt.pdf>
3. Dollar, C. (2000). Authentic Electronic Records: Strategies for Long-Term Access. Chicago: Cohasset Associates. (ISBN 0-9700640-0-4).
4. Work of Dr J. van Bogart for National Media Laboratory (NML) United States previously available online at <http://www.nml.org>. Site available through search-engine caches June 2001. Please note as the handbook goes to press a new publication, Data Storage Technology Assessment 2000 by Koichi Sadashige for the National Media

Laboratory and the National Technology Alliance will be available on CD from the NML.

5. PRO 1999. A Digital Preservation Strategy for the PRO. November 1999.
6. Woodyard, D. (1999). 'Practical Advice for Preserving Publications on Disk'. Presented at Information Online and Ondisc '99, Darling Harbour, Sydney, 21 January 1999.
<http://www.nla.gov.au/nla/staffpaper/woodyard2.html>

The Handbook was first compiled by Neil Beagrie and Maggie Jones and is now maintained and updated by the DPC.