Maintaining access to digital resources over the long-term involves interdependent strategies for preservation in the short to medium term based on safeguarding storage media, content and documentation, and computer software and hardware; and strategies for long-term preservation to address the issues of software and hardware obsolescence.This section is therefore divided into two parts: the first dealing with storage and maintenance of digital resources; and the second with strategies for their long-term preservation.

A preservation strategy for digital resources is most effective if it addresses the full life-cycle of the resource allowing the greatest efficiencies between data creation, preservation and use.This section should therefore be read in conjunction with related sections and chapters particularly the other sections of this chapter and **Media and Formats** .

Storage of digital resources supports both access and preservation. Depending on the needs of the organisation and the media, it may be necessary to create both preservation and access copies and to have strategies for each. We have used the term "digital preservation" in this handbook to define all the activities employed to ensure continued access to digital resources which have retained properties of authenticity, integrity and functionality.The term "archiving" can be substituted for preservation provided this definition remains. Archiving is usually interpreted within the computing industry simply to indicate that something has been stored and is no longer immediately accessible.The richer interpretation used here means that there will need to be more thought and preparation given to what resources are stored, how they are maintained and subsequently accessed and by whom.

There is no single definitive solution which can be applied for the preservation of any digital resource. However, an approach which is based on good management practices commenced as early as possible in the lifecycle of a resource, will safeguard the initial investment and facilitate authorised access at least for the short to medium term. Preventive preservation is as crucial a strategy in preservation programmes for digital resources as it is for non-digital material and good storage practice plays a major role in both. Key initial decisions needing to be made by institutions taking responsibility for short- or long-term preservation of digital resources will be:

1) Whether storage and/or preservation will be undertaken by the host institution or under contract to a trusted third party (see **Third Party Services** for discussion of issues relating to whether or not to outsource); 2) Which resources justify preservation and for what period of time.

The assumption in 2) is that not all resources can or need to be preserved forever, some will not need to be preserved at all, others will need to be preserved only for a defined period of time, a relatively small sub-set will need to be preserved indefinitely. Making this decision as early as possible will help to conserve resources for the most valuable digital assets.

This section deals with the range of strategies and approaches which will help to ensure important digital resources do not become inaccessible prematurely. Many constitute a relatively modest investment compared to the initial costs of creating the resource, which are often substantial.They can also represent significant cost savings longer term. In any event, failure to commit resources to managing digital resources throughout their lifecycle will inevitably result in their loss and/or costly restoration so investment in strategies to prevent this is eminently justified.

4.3.1 Storage and Maintenance

Storage media and file formats

General advice on storage media and file formats is provided in **Media and Formats** . Policy and selection of storage media and file formats will have implications for institutional strategies such as outreach and development of standards and best practice guidelines (see **Outreach** and **Standards and Best Practice Guidelines** **)** and for accessioning (see **Acquisition and Appraisal**). Decisions will need to be made during accessioning on whether to store resources as received or to reformat. A table outlining options, issues and requirements to assist with this decision process is provided in **Accessioning** .

Management of media and systems

Media refreshing and reformatting

Rationale

An essential management component for all digital media to avoid media degradation and to facilitate longer term preservation strategies.

Requirements

- Needs to be part of an ongoing regime so appropriate resources are required.
- Reformat data resources onto selected archival media if necessary.
- Write archive copies with different software to protect data against corruption from malfunctioning or virus- or bug-ridden software.
- Write archive to comparable magnetic media purchased from different suppliers to guard against faults introduced by the media's suppliers into their products or into batches of their products.
- Refresh or transfer archive copies to new media at specified times.This should take place:

- within the minimum time specified by the supplier for the media's viability under prevailing environmental conditions;
- when new storage devices are installed;
- when an audit discloses significant temporary or read "errors" in a data resource.

- Employ quality control procedure such as bit/byte or other checksum comparisons with originals to ensure the authenticity and integrity of items after media refreshing.
- Document actions taken when data resources are copied.
- Retain copies of the digital resource in its original format whenever some information or presentation of the resource may be lost or modified in reformatting.

Disaster recovery planning

Rationale

The development and use of a disaster recovery plan based on sound principles, endorsed by

senior management, and able to be activated by trained staff will greatly reduce the severity of the impact of disasters and incidents.

"The assumption is that with good disaster planning data recovery will be, under most circumstances, unnecessary. The problem is that while attention has been paid to disaster planning and the identification of good recovery procedures the effectiveness of these tend to depend upon pre-disaster effort. This effort often never takes place." (Waters and Garrett 1996)

Requirements

- Develop counter disaster plan to operate in the event of natural or man-made disasters. One model is the Disaster Recovery Procedures developed by the Data Archive, copied below, with the permission of the Data Archive.
- Ensure all relevant staff are trained in counter disaster procedures.
- Create archive copies of data resources at the time of their transfer to the institution.
- Store archive copies on industry standard digital tape or on other approved contemporary media.
- Store archive copies on and off site. Off-site copies should be stored at a safe distance from on-site copies to ensure they are unaffected by any natural or man-made disaster affecting the on-site copies.

Case study - disaster recovery procedures - Data Archive, University of Essex

The Data Archive is the UK national data centre for the Social Sciences funded by the Economic and Social Research Council (ESRC) and the Joint Information Systems Committee (JISC).The Data Archive has over 4000 mainstream digital datasets or studies, comprising over 125,000 individual files.

The digital storage system at the Data Archive is based on a Hierarchical Storage Management System (HSM) where the files appear to be local to the user but are mainly based on tape. As each file is requested it is either brought back from the disk cache on the system or automatically "restored" from the required tape. Any subsequent requests for that file are returned from disk cache.

Disaster recovery at the Data Archive is based around the resilience provided by creating multiple copies of the data and specified recovery procedures. Each file from any dataset has at least four copies and these are as follows:

Main copy This copy is held on the main area on the HSM file system.

Shadow copy At least one shadow copy is made. As files are updated, they are "shadowed" onto a separate tape in the main system. Multiple versions of these files are kept to allow staff to go back to a previous version of a file.

CD-ROM copy A CD-ROM is created for each dataset as part of the preservation procedure.This allows staff to access an alternate local source in the case of downtime of the main system and serves as an alternative long-term storage media. For each study all of the files are compressed and stored as a single zip file and written on to a CD-ROM. Subsequent updates to this study are created as complete zip files xxxx_2.zip and appended to the existing CD-ROM for that study.

Off-site near-line copy An off-site, near-line copy is kept in case of a major disaster at Essex. Due to restrictions of small file sizes on these systems, these are kept in the form of a range of datasets, which have been grouped together, compressed and encrypted.

Disasters can occur in different forms and at varying levels.The Data Archive has in place a range of recovery measures designed to meet any conceivable disaster.

- **Corrupt file**
  A file is supplied with corrupt information that is not detected through Data Processing
  Solution
  A. The file is re-requested from the supplier.
  B. Older version(s) of the file are retrieved from the shadow area and are either supplied back to the depositor or used to replace the corrupt file.
- **Unreadable file**
  A single file is unreadable from the media due to a bad block on a tape
  Solution

A. The tape is checked to make certain that this is an isolated problem. If it is found to affect the complete tape the corrupt media disaster recovery procedure is activated.

B. If the problem is isolated then the problematic file is recreated from the shadow area.

- **Corrupt media**

In this case a complete tape is damaged or cannot be reliably read.

Solution

A. If the tape was full and was set as read only and a refreshed tape was available then that could be copied to regenerate a new tape. B. If no retired refreshed media was available then a new tape could be created by retrieving the files from the shadow area, which are held on separate tapes.This process would require about 8 hours downtime of the HSM system.This process has been successfully used after a tape was damaged in the library due to a firmware fault on the DLT.

- **Corrupt shadow area as well as main area**

In this situation both the main and shadow areas cannot be read, nor any of the refreshed tapes.

Solution

This is very unlikely due to the number of checks that are made but in the event, the study or data would be re-created from the read-only CD-ROM copy. A CD-ROM copy is generated when the data is placed onto the preservation system and so would be up to date.

- **Complete loss of data at the University of Essex**

In this scenario, all of the data held at the University of Essex are unreadable and all of the systems are damaged beyond repair. (Major disaster.)

Solution

The main HSM systems would be built and data would be retrieved from the off-site holdings at ULCC.

Source:The Data Archive. Systems and Preservation Procedures(1999 unpublished) reproduced with the kind permission of the Data Archive.

Environmental conditions

Rationale

Appropriate environmental conditions will increase the longevity of digital storage media and help prevent accidental damage to a data resource or its documentation.

Requirement

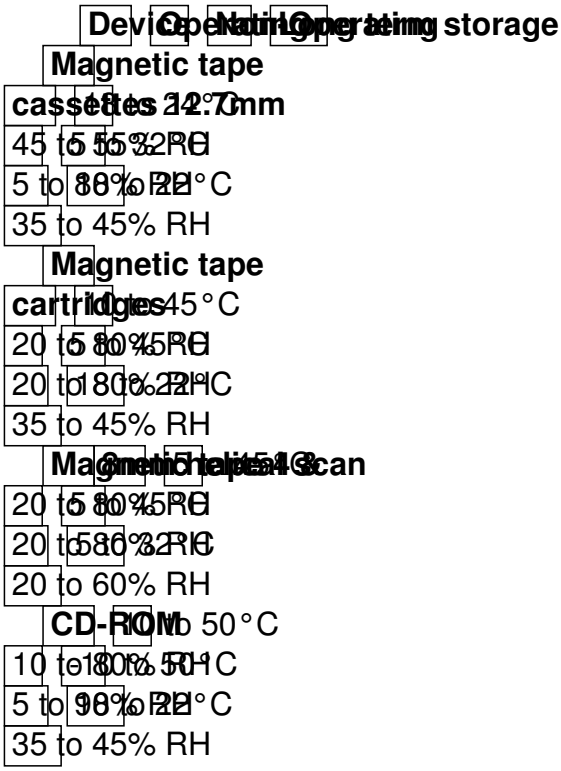
- Follow relevant guidance on environmental conditions for storage media in BS 4783.

Note: Most experts agree that large fluctuations in temperature and humidity are more damaging than having slightly higher than ideal temperature and Relative Humidity (RH). See, for example Van Bogart (1995) (DLM Forum 1997).

The following figure summarises British Standard 4783.

Figure 6

Summary of Environmental Conditions Recommended in BS 4783 for Data Storage Media



Extracts from BS 4783 reproduced with the permission of the British Standards Institution under licence number 2001/SK0280

- Establish guidance and procedures for acclimatising magnetic tape if moving between

significant variations in temperature (e.g. tapes moving from very cold external conditions should not be used before being acclimatised to warmer internal conditions).

- Establish procedures for monitoring environmental conditions.
- Minimise risk of damage from dust and other airborne pollutants.
- Prohibit smoking and eating in the storage area.
- Store away from direct sunlight.
- Provide additional protection in the form of enclosures for media.
- Provide storage facilities which minimise the threat from natural disasters such as fire and flood or to magnetic storage media from magnetic fields.
- Ensure any non-digital accompanying materials (e.g. codebooks, operating instructions) are also stored in appropriate environmental conditions.

Care and handling

Rationale

Appropriate care and handling will protect fragile digital media from damage.

Requirements

- Establish written guidelines and procedures based on available guidance (see **Further Reading** to this section and **Media and Formats**).

Audit

Rationale

There needs to be assurance that the resource has not been inadvertently or deliberately changed following refreshment and/or migration procedures and to check the readability and integrity of the data over time.

Requirements

- Check media periodically for their readability. Such checking may be conducted automatically in mass storage systems according to parameters set by system operators.
- Check the integrity of data files periodically using checksum procedures. Such procedures may be implemented automatically in mass storage systems according to parameters set by system operators.
- Employ appropriate security systems and procedures to protect the authenticity of items in your holdings (see **Security** below).

Security

Rationale

Rigorous security procedures will a) ensure compliance with any legal and regulatory requirements; b) protect digital resources from inadvertent or deliberate changes; c) provide an audit trail to satisfy accountability requirements; d) act as a deterrent to potential internal security breaches; e) protect the authenticity of digital resources; f) safeguard against theft or loss.

It is important to note that not all digital resources will require identical levels of security. Some, for example commercial in-confidence, will require more rigorous security regimes than less sensitive material. Guidance on levels of security can be found in BS 7799 Information Security Management (Tanner and Lomax-Smith 1999). All personal data will need to conform with the requirements of the Data Protection Act (1998) (
PRO 1999
).

Requirements

- Establish disaster recovery plan (see above).
- Control access to storage facilities and processing areas. Store in separate, preferably lockable area.
- Ensure no unauthorised access.
- Design audit features into mass storage systems and computerised physical access controls. Undertake regular random checks if automated audits are not feasible.
- Establish procedures to ensure no deliberate or inadvertent changes can take place.
- Ensure all legal requirements are met.
- Establish procedures for ensuring authenticity.
- Use passwords and user ids, and other network security procedures.
- Define system and area access privileges for staff.
- Assign specific staff responsibilities for data security and storage facilities.

Management of computer storage

Rationale

Unlike storage space for physical collections, computer storage is both reducing in cost and increasing in capacity all the time. Costs for processor capacity and storage media are expected to continue to drop (halving every 18 months at least according to Moore's Law) for several years to come (Kenney and Chapman 1996). However while storage is much less of a problem than it was, it conforms to good practice to establish policies and procedures which clarify what digital resources need to be accessible online, nearline or offline. Digital resources can be generated relatively easily, and the prospects for storage space to become cluttered with several versions of documents and other less valuable digital resources are quite high. It makes sense to establish when certain categories of resources may be automatically removed from online storage after a defined period of time, when others will be re-assessed, and which resources will be considered to be sacrosanct.

These decisions will need to be well documented and understood by all stakeholders within the institution.

Requirements

- Policies for maintaining documents on central file server (See **See Exemplars and Further Reading** , page 112, Storage and Maintenance, Oxford University Policy on Computer Archiving Services).

- Strategies for migrating to larger file server before full capacity is reached.
- Policies to identify which digital resources should be stored online.
- Retention policies to determine at what stage (if ever) online storage of digital resources will be re-assessed (see also **Acquisition and Appraisal** **).**

4.3.2 Preservation Strategies

This section is divided into primary preservation strategies and secondary preservation strategies. Primary preservation strategies as defined here are those which might be selected by an archiving repository for medium to long-term preservation of digital materials for which they have accepted preservation responsibility. Secondary preservation strategies are those which might be employed in the short to medium term either by the repository with long-term preservation responsibility and/or by those with a more transient interest in the materials. Chronologically, secondary strategies may precede primary strategies. Some secondary strategies may substantially defer the need for, or alternatively greatly strengthen, primary preservation strategies so describing them as secondary strategies does not necessarily imply their inferiority.Two strategies dominate current options for preserving digital resources long-term, these are migration and emulation. Both have champions and detractors, both have acknowledged difficulties. The need for both may also be deferred and/or simplified if appropriate preventive preservation procedures such as storage and maintenance (see **Storage and Maintenance**) and selected secondary preservation strategies, have been used.

The other potential long-term strategy,

to an analogue preservation format, differs from the other strategies in two important ways:

1. It can only sensibly be considered for a relatively small category of digital resources and is patently inappropriate for the increasing numbers of more complex digital resources being created.
2. By its nature, it loses the digital characteristics of the resources it converts and is therefore a preservation strategy for some digital resources, as opposed to a digital preservation strategy, where the essential aim is to retain the digital characteristics of the resource. The latter should be preferred.

Another option represented here as a secondary strategy is digital archaeology (secondary strategy 7).This is not precisely a preservation strategy at all but rather when the absence of preservation strategies has left valuable resources inaccessible.

It should be emphasised that employing a judicious mix of secondary strategies 1-5 combined with responsible storage and maintenance decisions in **Acquisition and Appraisal** has the potential significantly to reduce both risks of losing access to digital resources in the short-term and costs of preserving access to them in the long-term.

Primary preservation strategies

Preservation strategies selected by archiving repositories with long-term preservation responsibility for the digital materials in their care. It should be noted that discussion of costs in this context is of necessity based on educated assumptions as opposed to empirical evidence gathered over a very long timeframe. Cost models for complex digital materials particularly those of recent origin are still at the research stage at the time of writing.

Migration

Description

A means of overcoming technological obsolescence by transferring digital resources from one hardware/software generation to the next.The purpose of migration is to preserve the intellectual content of digital objects and to retain the ability for clients to retrieve, display, and otherwise use them in the face of constantly changing technology. Migration differs from the refreshing of storage media in that it is not always possible to make an exact digital copy or replicate original features and appearance and still maintain the compatibility of the resource with the new generation of technology.

(Note: There are differing degrees of migration, ranging from relatively straightforward conversion to a major paradigm shift. Obviously the latter category will be most relevant to the disadvantages outlined below. It should also be noted that by using the secondary preservation

strategy of standards, it may be possible to delay the need for migration).

Advantages

- Procedures for simple migration are well established.
- Is currently the preferred strategy for most digital archives.
- May become simpler as technology advances and range of platforms diminishes.
- A recent CLIR publication has produced a risk assessment tool to assist decision-making (PRO 1999).

Disadvantages

- Cost - requires special program to be written for complex migrations.
- Can be time-consuming and complex.
- Likely to lose some functionality and look and feel of original.
- May compromise the integrity of the originals unless stringent quality control procedures to ensure authenticity are in place.
- More complex digital resources may be migrated with significant loss of functionality.
- Needs to occur at regular intervals throughout the life of the resource. See Rothenberg (see note) for more detailed discussion of what he considers to be major drawbacks to migration as a digital preservation strategy.

Requirements

- Written policies and guidelines, including selection policy for materials to be migrated.
- Quality control procedures.
- Rigorous documentation of migration procedure.
- Preservation metadata and documentation (see **Metadata and Documentation** **).**
- Migrate data whenever there is a software upgrade or a new software application is installed.
- Ensure the migration results in little or no loss in content or context.
- Employ strict quality control procedures that may include testing the migration programme with a sample of records or bit/byte or checksum comparisons of migrated and original data.

- Retain copies of the digital resource in its original format whenever some information or presentation of the resource may be lost or modified in migration.

Related strategies

- Storage and maintenance.
- Backwards compatibility.
- Permanent identifier.
- Validation procedures.
- Conversion to standard formats.

Emulation

Description

A means of overcoming technological obsolescence of hardware and software by developing techniques for imitating obsolete systems on future generations of computers.

Advantages

- Recreates the functionality, look and feel of the original.
- Avoids repeated costs associated with migration (though see also disadvantages below).
- May offer the best prospects for more complex digital resources.

Disadvantages

- Is still in the research stage and requires further practical testing (see CAMiLEON project (see note) and Rothenberg (see note

), (
PRO 2000
). See also Bearman (
PRO 1999
) (1999) for a critique of emulation as a viable preservation strategy).

- May only be able to emulate part of the functionality, look and feel of the original.
- Is likely to be very costly unless it has economies of scale. New emulators need to be built for major computer paradigm shifts; it is possible that these costs may even exceed the savings of repeated migration costs.
- Software copyright issues need to be addressed and may be extremely complex.
- There must be rigorous documentation of hardware and software requirements.These have rarely been documented to this level of detail in the past and would require concerted effort and resources.

Requirements

- Appropriate storage and maintenance procedures (see **Storage and Maintenance**).
- Written policies and guidelines.
- Preservation metadata (see **Metadata and Documentation** **).**
- Detailed documentation on hardware and software specifications.

Related strategies

- Storage and maintenance.
- Encapsulation.
- Permanent identifiers

Secondary preservation strategies

Secondary preservation strategies are those which might be selected either by the archiving repository with long-term responsibility for the preservation of digital materials and/or by those with a more transient interest in the digital materials they have created and/or acquired. A judicious combination of secondary strategies and appropriate storage and maintenance (see
Storage and Maintenance
) can be a cost-effective means of ensuring continued access to digital materials for as long as

they are needed, either deferring or in some cases, even avoiding, the need for primary preservation strategies.

Technology preservation

Description

A means of overcoming technological obsolescence by retaining the hardware and software used to access the digital resource. It should be noted that the current definition of this strategy involves individual institutions needing to maintain both hardware and software for all materials they create and/or acquire. A variation of this strategy has been suggested which involves the setting up of a facility offering documentation for hardware and software and file format specification (PRO 1999), (DLM Forum 1997). If these recommendations were implemented, this variation on the technology preservation strategy could become a much more feasible proposition and provide valuable support for genuinely long-term emulation or migration strategies.

Advantages

- Storage retains the functionality, look and feel of the original.
- Storage delays the time when other preservation strategies are required.
- Storage may be the most practical interim strategy for complex digital resources.

Disadvantages

- Can only be used as a short- to medium-term strategy. Is not viable long-term as defined here.
- Technical support will inevitably disappear within a relatively short timeframe.
- Facilitating access will become increasingly problematic over time.

Requirements

- Policies and guidelines regarding access.
- Documentation of hardware and software maintained.
- Metadata required to maintain the hardware and software.

Related strategies

- Storage and maintenance.
- Conversion to standard formats.
- Backwards compatibility.
- Adherence to standards.

Adherence to standards

Description

Adhering to stable and widely adopted open standards when creating and archiving digital resources.These are not tied to specific hardware/software platforms and thus can defer inaccessibility of digital resource due to technological obsolescence. Can either be self-imposed by institutions creating digital resources, or imposed by agencies receiving digital resources (see also **Standards and Best Practice Guidelines** and **Media and Formats**).

Advantages

- Using stable open standards will delay the time when more costly strategies are needed.
- Using stable standards will reduce the complexity, and therefore costs, of longer-term preservation strategies.
- Can simplify migration and achieve economies of scale in migrating similar items.
- Can benefit creators as well as long-term preservation. Helps to distribute some of the effort over the lifecycle of resources.

Disadvantages

- Dependent on creators being able and/or willing to comply or later conversion by the archive.
- Stable standards are not available for some formats.
- Even when stable standards do exist, they are themselves subject to inevitable change as they evolve into new versions.
- Proprietary extensions are relatively common and generally not as well documented as the standard itself.

Requirements

- Knowledge of all relevant standards for all categories of digital resources acquired by the institution.
- Written guidelines on preferred and acceptable standards.
- Institutional strategies for outreach, collaboration, standards and best practice.
- Technology watch on standards activities.

Related strategies

- Adherence to standards will facilitate all other digital preservation strategies.

Backwards compatibility

Description
Being able to retain accessibility to a digital resource following upgrade to new software and/or operating systems.

Advantages

- Defers for a period the need for primary preservation strategies.
- Is being offered by increasing number of vendors.

Disadvantages

- Is not routinely offered by all vendors.
- Can only be of short- to medium-term value.
- Even when it exists it cannot be expected to last indefinitely.
- Its continued availability is dependent on market forces which are notoriously volatile. It may therefore cease to be available with little or no warning.

Related strategies

- Storage and maintenance.

Encapsulation

Description

Grouping together a digital resource and whatever is necessary to maintain access to it.This can include metadata, software viewers, and discrete files forming the digital resource.

Advantages

- Ensures all supporting information required for access is maintained as one entity.
- Can potentially overcome some of the major disadvantages of alternative strategies.
- Provides a useful means of focussing attention on what elements are needed for access.

Disadvantages

- Can produce very large files with duplication (e.g. of viewers) across the collection unless these links are maintained.
- Encapsulated software is still open to rapid technological obsolescence.

Related strategies

- Emulation

Permanent identifiers
Description
A means of locating a digital object even when its location changes. Examples are Universal Resource Names (URN's); Handles; Digital Object Identifiers (DOI's); Persistent Uniform Resource Locators (PURLs)

Advantages

- Critically important in helping to establish the authenticity of a resource.
- Provides access to a resource even if its location changes.
- Overcomes the problems caused by the impermanent nature of URLs.
- Allows interoperability between collections.

Disadvantages

- There is no single system accepted by all.
- The costs of establishing or using a resolver service.
- Is dependent on ongoing maintenance of the permanent identifier system.

Related strategies

All, except Conversion to Analogue Formats.

Converting to stable analogue format

Description

Converting certain valuable digital resources to a stable analogue medium such as permanent paper or preservation microfilm or, more recently, nickel disk readable by electron microscope.This cannot be recommended as more than a pragmatic interim strategy for a small category of digital materials, pending the development of more appropriate digital preservation strategies.

Advantages

- Is no longer vulnerable to technological obsolescence assuming preservation quality microfilm or permanent paper is used.
- Should essentially be a "once only" cost for conversion.
- Will guarantee accessibility for hundreds of years provided it is converted to an archival standard and stored in archival conditions.
- May be a pragmatic interim strategy pending the development of infrastructure for more appropriate digital preservation strategies.

Disadvantages

- Loses functionality of original digital resource.
- Can only sensibly be considered as an option for digital resources which do not utilise or require the full functionality of digital technology.
- Has already caused difficulties even when used for simple text emails[6].
- Cannot be considered for more complex digital resources where loss of functionality would at best diminish, if not destroy, the usefulness and integrity of the resource.
- Loses the advantages of digital technology, for example the convenience of use, and efficient use of space.
- Costs of conversion to archival standard and storage in archival conditions (the latter cost will be recurrent and the cumulative cost will be significant over time).

Requirements

- Policies and guidelines clearly documenting rationale for adopting strategy and category of resources it may be used for.

Related strategies

- None, this is not a digital preservation strategy but a mechanism to preserve the information content of certain digital resources.

Digital archaeology

Description

Rescuing digital resources which have become inaccessible as a result of technological obsolescence and/or media degradation. Not so much a strategy in itself as a substitute for one when digital materials have fallen outside a systematic preservation programme.

Advantages

- There are a growing number of specialist third party services offering this service.
- It has been shown to be technically possible to recover a wide range of information from damaged or obsolete media (though not necessarily in the same form).

Disadvantages

- Much more costly long-term than bona fide digital preservation strategies.
- Is unlikely to be cost-effective for anything other than the most highly valued digital resources.
- Potentially useful materials which do not justify the costs involved will be lost.
- Risk that some digital materials may not be able to be successfully rescued.
- Poor management of initial investment.

See Exemplars and Further Reading