

## 4. Organisational Activities

### 4. Outline

#### Intended primary audience

Creators and publishers of digital resources, third-party service providers, operational managers and staff with responsibility for implementing institutional activities of relevance to digital preservation. It is assumed that these will include a) staff from structurally separate parts of the organisation, and b) a wide range of knowledge of digital preservation, from novice to sophisticated; c) both technical and non technical perspectives; d) a wide range of functional activities with a direct or indirect link to digital preservation activities.

#### Assumed level of knowledge of digital preservation

Wide-ranging, from novice to advanced.

#### Purpose

- To provide pointers to sources of advice and guidance aimed at encouraging good practice in creating and managing digital materials. The importance of the creator in facilitating digital preservation is stressed throughout the handbook but particularly in **Creating Digital Materials**. Good practice in digitisation and other digital materials creation is crucial to the continued viability of digital materials.
- To raise awareness of factors which need to be considered when creating or acquiring digital materials.
- To provide pointers to helpful sources of advice and guidance for both novices and those who have already begun to think through the implications of digital

### 4.1 Creating Digital Materials

"The first line of defense against loss of valuable digital information rests with the creators, providers and owners of digital information." (**Waters and Garrett 1996**)

The Task Force on Archiving of Digital Information articulated one of the earliest acknowledgements of the crucial role of the creator in helping to ensure long-term access to the digital resources they create. This view has been reiterated in many other documents since the Final Report of the Task Force was published. Clearly, most individual creators cannot be expected to take on long-term commitment to preserving the digital content they create (**DLM Forum 1997**). Every digital resource has a life cycle and different stakeholders and interests within this. However, it is both achievable and highly desirable that a dialogue is established between them when issues of long-term preservation are involved. Given the crucial role of the creator in undertaking short to medium-term preservation and at least facilitating medium

to long-term preservation, this section will focus on encouraging good practice in creation of digital materials which will assist in their future management.

A major source of current activity and investment is in the digitisation of analogue materials, in particular digital imaging. There are many useful tools which provide assistance with various stages of digitisation projects. This section of the handbook will not attempt to duplicate work done by others by producing a detailed decision tree for digitisation but will act as a means of flagging issues relevant to the management of digital objects and provide links to more detailed sources of advice and guidance.

There is also a wide range of digitisation methods and this section is not intended as a digitisation guide or manual for different methods of capturing information. Our focus remains the implications for digital preservation in the creation process. Two areas have been selected, creating digital surrogates and creating electronic records as being of most widespread interest and illustrating general preservation principles for other data creation methods.

The emphasis on digitisation in this section reflects its current importance as increasing numbers of institutions embark on digitising parts of their collections. It is important to reinforce that this handbook is not considering the potential of digitisation as a preservation reformatting tool. The emphasis throughout the handbook is on the preservation of "born digital" materials, or the products of digitisation (the digital surrogates themselves), not the preservation of the analogue originals.

Many digitisation projects cite enhanced access as the major objective, a perfectly legitimate objective but unless due care and attention is given to how that access can be maintained over time, it may well be short-lived. This section of the handbook makes the assumption that it is highly unlikely that all current digitisation initiatives are being undertaken with due regard to the long-term viability of the digital surrogates they are creating. A related assumption is that it would be useful to encourage good practice in creating digital materials and to point to existing sources of guidance.

A second major source of current activity is in the creation of electronic records. This section is divided into two, the first focusing on the creation of digital surrogates through digitisation and the second on the creation of electronic records. Both have excellent sources of advice and guidance and key references are provided in an annotated reading list following the section.

#### 4.1.1 Creating Digital Surrogates

The following diagram (Figure 3) illustrates how the relationships between the various elements should ideally flow within an institution. For the sake of simplicity, the diagram looks at the broad issues as they apply to long-term preservation, referring to more detailed guidance documents, as appropriate. It suggests that a strong corporate presence, in the form of policies and associated strategies, is required in order to provide the necessary guidance and authority to staff involved in institutional digitisation projects. Consideration of how the digital surrogates will be maintained needs to be made as early as possible, preferably at the

84

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

design stage. It is also important to note that the broad model applies to all activities, not just digitisation, and the Further Reading section reflects this wider perspective.

**Figure 3**

Creating Digital Surrogates - Management Context and Checklist

**Corporate Policies**

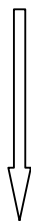
Outlining broad policy towards selection for digitisation, purposes of digitisation; management of digital surrogates etc. Should also indicate responsibility.

See Further Reading for models and frameworks and checklist for elements which may need to be addressed in policy.



**Corporate Strategies**

Strategies to put into practice the principles articulated in corporate policy statements. See Further Reading for models and frameworks and checklist below for issues needing to be considered.



**Corporate Procedures**

Guidelines for operational activities which are clearly linked to corporate strategies. See checklist below for elements which need to be considered.

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

Checklist (see Reference 3)		Issues in Preserving Digital Surrogates
1	Assessment of need for digitisation	Has the material already been digitised? If so, is it to an appropriate standard and readily accessible?
2	Finding funds for the project	What archiving policies exist, both from the funding agency (if externally funded) and the institution with prime responsibility for the project?
3	Planning the project and assigning resources	Need to set aside recurrent funds for maintenance of the digital copies as well as one-off funds for conversion. Ensure all relevant stakeholders are aware of the project (for example, if another part of the organisation or an external agency is expected to maintain the resource, they will need to be included in discussions at this point, if not before)
4	Selection of materials	Copyright. Need to ensure permission is given both to digitise the original and to make copies of the digital copy for the purposes of preservation. For further information, see <b>Legal Issues</b> and <b>Rights Management</b> . Condition and completeness of original. Is it capable of being re-scanned at a later date if the digital copy is lost?
5	Decide how the information content needs to be organised (for example, searchable text databases and/or document page images)	Selection of appropriate file formats and storage media for both master/archive copies and derivatives.
6	Decide digitisation method appropriate to analogue original and goals of the project.	Details of the digitisation method need to be documented and attached to the metadata record to enable future management.

Checklist (see Reference 3)		Issues in Preserving Digital Surrogates
7	Preparing originals for digitisation	Documentation. Will the originals be retained? (if scanning paper records, the PRO advises not to take any action on discarding the originals until it is established that a) the electronic version is legally admissible and/or b) the electronic version is capable of long-term preservation ( <b>PRO 1999</b> .) For collection material, Kenney and Chapman provide a decision tree for deciding whether or not to retain the originals post-digitisation ( <b>Kenney and Chapman 1996</b> ). The latter will of course not be an issue for projects digitising valuable treasures within a collection, the main issue then will be whether or not the original is too fragile to be re-scanned at a later date if the digital copy is lost. In any of these cases, if the digital copy becomes the primary means of access, it will be subject to the same requirements as born digital material.
8	Conversion	Documentation of technical characteristics. Compression algorithm (if used); bit depth required; scanning resolution etc. Create backup copies as soon as conversion is undertaken.
9	Quality assurance checks	Digital surrogate needs to be of an acceptable preservation quality. If using third party services, need to ensure documentation clarifies responsibility for quality assurance.
10	Final indexing and cataloguing	Metadata for resource discovery and for managing and preservation of digital copy.
11	Loading data into computer systems	Document storage requirements for access and preservation copies (if different). Make backup copies as appropriate. Note: the PRO recommends a minimum of four copies as a general rule, with a minimum of two on separate storage media. If the analogue original is in good condition and capable of being re-scanned in case of loss of the digital copy, two copies, each stored geographically separately, may be sufficient ( <b>PRO 1999</b> ).

Checklist (see Reference 3)		Issues in Preserving Digital Surrogates
12	Implementing archiving and preservation strategies or transferring to a preservation agency	Required standards for formats, storage media, documentation, and transfer procedures. Storage of masters and backup copies. Strategies for media refreshment and changes in technological environment.

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

#### 4.1.2 Creating Electronic Records

Most records created during the day-to-day work that takes place in the public and private sector are now created electronically. Increasingly, they will also be distributed and accessed electronically. In the UK, the recent white paper, Modernising Government (**see Reference 7**) aims to have all newly created public records electronically stored and retrieved by 2004 and a strategy has now been developed to achieve this aim (**see Reference 8**). Combined with recent legislation on Data Protection and Freedom of Information (**see Reference 8**), this makes efficient and effective management of electronic records of pressing concern. Records management is a key aspect of the Freedom of Information Bill and a code of practice to be issued by the Lord Chancellor will underpin the Act (**PRO 2000**). In order to fulfil their legal and accountability responsibilities, organisations in the UK will need to ensure they plan for efficient and effective electronic records management (ERM). There are also organisational implications (See also Preservation Issues and **Organisational Activities** for more discussion of these issues).

Three major sources of guidance have been consulted for this section. They are:

- Electronic Records Management: Framework for Information Age Government. March 2000.  
<http://www.e-envoy.gov.uk/assetRoot/04/00/22/94/04002294.rtf>  
**Update 26 Nov 2004**  
Electronic Records Management: Framework for Information Age Government. April 2000.  
<http://e-government.cabinetoffice.gov.uk/assetRoot/04/00/22/94/04002294.rtf>  
**Update 12 Mar 2007**  
Replaced with link to PDF  
[http://archive.cabinetoffice.gov.uk/e-envoy/resources-pdfs/\\$file/erm.pdf](http://archive.cabinetoffice.gov.uk/e-envoy/resources-pdfs/$file/erm.pdf)
- DLM Forum. (1997). Guidelines on the Best Practices 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/>
- Public Record Office. (1999). Guidelines on the Management, Appraisal and Preservation of Electronic Records. Volumes 1 (Principles) and 2 (Procedures). Kew: Public Record Office.  
<http://www.nationalarchives.gov.uk/electronicrecords/advice/guidelines.htm>



There is a commonality of themes in these three sources and the following checklist draws heavily on all three.

### Formal corporate policies.

"It needs to be clearly understood across the department that everyone is responsible in some way for records and that responsible behaviour is implanted throughout all relevant operational activities. Establishment of a defining framework of formal corporate policies on electronic records is one principal means of helping to achieve this goal." (**PRO 1999**)

As in the previous section on digitising analogue collections, it is difficult to overemphasise the importance of corporate policies promulgated throughout the organisation and re-visited at regular intervals to ensure continuing relevance in a rapidly changing environment. The Public Record Office Guidelines also reinforce the need to ensure that the policy does more than pay "lip service" to accountability requirements:

"Most important of all, a corporate policy must be agreed to. Writing the words of a policy statement is much less difficult than the process of gaining agreement to them ...a policy which is ignored is worse than no policy at all." (**PRO 1999**)

The policy should address issues such as:

- The status of electronic records within the organisation and broad definitions of what they constitute.
- Broad definition of records which constitute permanent records worthy of long-term preservation.
- Whether long-term management will be undertaken by a third party service provider or in-house.
- Roles and responsibilities within the organisation.

### Corporate strategies

The best way to optimise the management of electronic information is to define a coherent global strategy from the outset, ensuring that everyone concerned is involved. One solution is to set up a multidisciplinary team to define and monitor the strategy (**DLM Forum 1997**).

The suggestion above recognises the fact that there needs to be a mechanism which will bring together the range of expertise necessary to develop effective strategies. The implicit assumption in monitoring the strategy is also that there will be clearly defined timeframes with achievable targets to monitor.

The following issues should be addressed in corporate strategies and may well require other supporting documents setting out in more details how the strategies can be achieved:

- Authenticity - What organisational and technical strategies will ensure that the electronic record is reliable and legally admissible?
- Appraisal and retention periods - How will they be developed and applied?

- Migrating records worthy of permanent preservation to new systems - How will they remain accessible and usable for as long as they are needed?
- Selection of media and formats - What standards should be used for various categories of electronic documents across the organisation?
- Metadata - How will essential metadata be defined and how can it be ensured that it will remain linked to the corporate record?
- Training and awareness raising for staff - What ongoing training requirements and supporting guidelines are required to support good records management practice?
- Compliance with the corporate policy - How will compliance with the corporate policy be monitored?

### Corporate procedures

Procedures developed by the Records and IT Manager need to define:

- The design of the recordkeeping system.
- The types of electronic records which need to be captured by the system.
- What documents need to be captured as records.
- What naming conventions should be used.
- What metadata needs to be kept with the records.

### Training and guidance

"New record keeping skills are required in a fully electronic environment of end-users as creators and users of records. They will have more responsibility for correctly identifying and dealing with electronic records at the point of creation, and these shifts imply significant cultural change in attitudes and behaviour towards record-making and use." (see **Reference 13**)

Most organisations will need to undergo a cultural shift which places more responsibility on the creators of records than has previously been the case. Staff will need to be aware of the following, through guidelines and training:

- At what point does the document they are working on become a formal record and therefore managed as part of the Electronic Records Management of the organisation?
- What metadata should they provide?
- What naming conventions should they use?

See **Exemplars and Further Reading**

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

## 4.2 Acquisition and Appraisal, Retention and Review

### 4.2.1 Appraisal and Selection

In a digital environment, decisions taken at creation and selection have significant implications for preservation. The link between access and preservation is far more explicit than for paper and other traditional materials, as access to a digital resource can be lost within a relatively brief period of time if active steps are not taken to maintain (i.e. preserve) it from the beginning. As the interactive **Decision Tree** indicates, if it is neither feasible nor desirable to preserve a digital resource across various changes in technology, then its acquisition should be re-evaluated. While many of the same principles from the traditional preservation environment can usefully be applied, policies and procedures will need to be adapted to the digital environment. In a print environment, the decision to select, and the decision to preserve, can be taken quite separately and within a timeframe which may span several decades. The brief period during which digital resources will inevitably become inaccessible means that it makes sense to make decisions about selection and preservation simultaneously.

While this may mean that greater rigour is required in selecting digital resources than for printed or other analogue material, it will avoid costs which will otherwise occur further down the track as retrospective preservation of digital resources is not recommended. In these cases, as the digital resources become inaccessible, the only pragmatic option is likely to be to de-select, an activity which is in any case not without cost and one which should preferably occur as a result of a conscious decision based on sound policies, rather than by default.

Accurate documentation is also crucial in the digital environment. This will provide not only essential details for managing the resource over time but also information on context without which there may be little point in preserving the digital object itself even if it is technically feasible to do so. In the accompanying **Decision Tree**, it is suggested that acquisition be re-evaluated if documentation is inadequate.

In the case of networked digital resources, where providing access to a resource does not necessarily require bringing the resource physically into a collection, the concept of acquisition is quite different from traditional collections. There are a range of options available to provide access or to build 'virtual collections'. For example, making copies/mirrors for access, providing a hyper link to a resource, online catalogues and finding aids.

Whereas acquiring a resource in the non-networked environment usually implies keeping it, in the networked digital environment, it is possible to provide access to a resource without undertaking any preservation commitment either short or long-term. A number of institutions have adopted a selection policy based on levels of acquisition. For example, the Berkeley Digital Library sunsite (**DLM Forum 1997**) adopts four levels (Archived; Served; Mirrored; Linked), the AHDS (**Tanner and Lomax-Smith 1999**) have articulated five levels (Archived; Served; Brokered; Linked; Finding Aids); while the National Library of Canada (**PRO 1999**) has three levels (Archived; Served; Linked). Adapting policies to the digital environment in examples such as these is likely to be the most cost-effective means of ensuring appropriate management and continued access to the most important digital resources. In practice

adopting collection levels and being explicit about preservation commitments is a crucial element of preservation policy and procedures.

In a digital environment acceptance of preservation responsibility implies significant costs. On the other hand, failing to consider short and long-term access at selection is likely to mean loss of the resource shortly after acquisition. In some cases (potentially many cases for electronic publications), an institution may be reluctant to take primary preservation responsibility for materials it acquires if it feels that interest in its preservation is so widely shared that it would constitute an unfair burden on their own institution. This emphasises the need for collaboration between institutions and the need to establish equitable agreements for shared efforts where necessary. The accompanying **decision tree** for appraisal and selection is based on the assumption that the resource has not yet been acquired and indicates a number of points at which cost implications will need to be taken into account before the decision to proceed with acquisition. It suggests that, at these points, difficult decisions may need to be made about whether the resource justifies the costs required or whether it is preferable not to proceed with acquisition.

See **Exemplars and Further Reading**

#### 4.2.2 Retention and Review

Selection for long-term retention will normally occur at acquisition but can be an iterative process occurring at later stages once an item is already in the collections. The term retention and review is used here for this iterative process. The decision process mirrors steps included in the **decision tree** in the previous section and the tree can be adapted for this purpose.

Employing evaluation criteria and selection procedures for all potential digital acquisitions ensures that collections development is carefully prioritised and sustainable. The use of such criteria and procedures should minimise the frequency and need for retention and review decisions as acquisitions are carefully evaluated and justified prior to entering the collections. Digital items acquired over time and before institutional policies and procedures were in place will normally require such review. This may be one of the first steps that an institution undertakes in implementing a digital preservation policy: quantifying its current digital holdings and assessing preservation risks (see **Technology Watch**).

Archives use the series concept for a body of records that share similar characteristics. Typically, many series are on-going for decades. However, the scope and coverage of a digital series may change over time and certainly technology considerations are likely to change and some attention must be given to a careful evaluation as each accession is transferred to the archives.

Over time the need may also arise to review collections and collections policy to reflect changing needs and circumstances. The necessity of making early decisions on selection for preservation in a digital environment (without the period of hindsight which is often available in analogue environment) may mean that future review may be necessary in the preservation life cycle of electronic resources.

In a digital library environment where collection levels have been employed, digital resources in any collection level category can be subject to periodic review, re-designated from one level to another, withdrawn, or de-accessioned as required to meet changing needs and circumstances. However, for items selected for permanent preservation it is anticipated that review and de-accessioning will occur in rare and strictly controlled circumstances. For other collection levels such as mirrored or licensed resources review criteria may include:

- A sustained fall of usage to below acceptable levels.
- The availability of content elsewhere to a higher degree of quality or at considerably lower cost.

Content that has been superseded or is no longer sufficiently accurate to justify maintenance in active form. In such cases, the content may be retained together with subsequent editions or withdrawn.

- Expiry or termination of a licence or data exchange agreement and withdrawal/return of a digital resource to the data supplier.
- Cost to sustain the data resource outweighs the value/benefit received.
- Deterioration in the quality service provided by a supplier or deterioration in the accessibility of content due to poor updating of indexing, imaging, or other characteristics internal to the data resource.

Within archives and records management professions the use of retention periods and schedules is well established. Records may be destroyed at the end of their retention period, retained for a further period, or transferred to an institution for long-term preservation.

In any collection environment it is important that written procedures are in place for the process of retention and review. The timescales, circumstances, and authorisation procedures for the review should be clearly stated. Depending on the institution's business environment, its users and depositors may be consulted as part of the process. Any recommendations may then be referred for approval to management and committees as appropriate to the size and significance of the resource.

Where a recommendation is made to de-accession an archived resource there should be procedures to consult with other stakeholders to determine whether transfer to another organisation should occur. In such cases the institution should agree conditions of transfer which include acceptable levels of care for the resource and access to it as appropriate for educational and research users.

Accessioned digital resources that have not been retained after review should retain their entry in any institutional catalogue with comments identifying the process undertaken and any transfer details.

### 4.2.3 Accessioning

Institutions should develop a range of accessioning procedures which support their preservation policies and objectives. These may include elements from the following list as appropriate to the item being accessioned:

#### Transfer procedures and guidelines

Most institutions will need to develop procedures and documents to support the smooth transfer of digital resources from suppliers into their collections. Figure 5 below outlines options for transfer and accessioning of file formats and storage media. Decisions on file formats and media (see Storage and Preservation) will support and be interdependent with this process.

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

**Figure 5**

Options for Transfer and Accessioning of File Formats and Storage Media

Options	Issue	Requirements
		<ul style="list-style-type: none"> <li>• Policy on storage formats.</li> <li>• Technology Watch on developments in storage formats.</li> </ul> <p>(all options)</p>
<p>Limit range of file formats received</p> <p>Limit range of media received (most cost-effective long-term option)</p>	<ul style="list-style-type: none"> <li>• Simplifies management and reduces overall costs.</li> <li>• Depositor may lack resource or expertise to comply.</li> <li>• Wide variety of file formats used and proprietary extensions to open standards.</li> <li>• Media used for transfer potentially can be used for long-term storage.</li> </ul>	<ul style="list-style-type: none"> <li>• Guidelines on preferred formats.</li> <li>• Degree of influence over the deposit.</li> <li>• Outreach and collaboration strategies to achieve desired outcomes.</li> <li>• Guidelines on preferred transfer media and transfer procedures.</li> </ul>
<p>Accept as received but convert to standard file format</p> <p>Accept as received but convert to standard storage media format</p>	<ul style="list-style-type: none"> <li>• Simplifies management and reduces longer term costs.</li> <li>• May not be technically feasible to convert to standard format.</li> <li>• It will be necessary to check that accidental loss of data has not occurred.</li> </ul>	<ul style="list-style-type: none"> <li>• Copyright permissions or statutory preservation rights.</li> <li>• Resources and technical expertise at host institution.</li> <li>• Election of preferred formats.</li> <li>• Documentation of native formats to allow conversion.</li> <li>• Integrity checks for conversion process.</li> </ul>

Options	Issue	Requirements
<p>Accept and store as received (least cost-effective option long-term, despite lower initial costs)</p>	<ul style="list-style-type: none"> <li>• Complicates management and increases costs of managing resources over time.</li> <li>• High risk option, particularly if large numbers of digital resources are being collected.</li> <li>• A choice of file formats may be available. That deposited may not be the most suitable for preservation.</li> <li>• Storage media may be of unknown quality and suitability for long-term preservation.</li> <li>• Formats may be obsolete or not supported within the institution.</li> </ul>	<ul style="list-style-type: none"> <li>• Clearly defined priorities for both short and long-term preservation.</li> <li>• Ability to address issues such as encryption, proprietary software etc. in received items.</li> <li>• Ability to ensure future access to information contained in the item.</li> </ul>

## Procedures to prepare data and documentation for storage and preservation

### **Unique numbering**

Each data resource accessioned by an institution should be allocated a unique identifier. This number will identify the resource in the Institution's catalogue and be used to locate or identify physical media and documentation. In the event of a resource being de-accessioned for any reason, this unique number should not be re-allocated.

### **Preferred marking and labelling**

At a minimum all physical media and hard copy documentation should be marked with the unique number allocated to the resource, and any additional information required by the institution easily to identify content and formats.

### **Handling guidelines**

Handling guidelines for accessioning staff should be developed reflecting storage and preservation staff advice on best practice for different media (see **Media and Formats**).



## Validation

Validation checks should be carried out by the institution on the transfer media, content and structure of deposited data resources, and on any accompanying documentation. Validation procedures may be adapted in the light of the volumes of material and resources available in the acquisitions section. It may be possible to automate some of the validation procedures but others can only be undertaken manually. Such checks may include:

- Scanning for computer viruses.
- Checking media and files can be read.
- Checking completeness and accuracy of paper based or digital documentation.
- Checking description and intellectual content of the resource.
- Checking structure and formatting of the resource.
- Procedures for documenting validation checks and any discrepancies encountered.
- Procedures for checking and if possible resolving discrepancies with the supplier.

## Re-formatting file formats

Where the file formats used to transfer the resource are unsuitable for long-term preservation, the Institution may re-format the resource onto its preferred file formats. In addition to archive formats, versions in other formats suitable for delivery to users may also be produced from the original (see **Storage and Preservation**).

## Re-formatting storage media

Where the storage media used to transfer the resource are unsuitable for long-term preservation, the Institution may re-format the resource onto its preferred media (see **Storage and Preservation**).

## Copying

Multiple backup copies of an item may be generated during accessioning as part of institutions' storage and preservation policy and to enable disaster recovery procedures (see **Storage and Preservation**).

## Security

System and physical security policies and procedures should be in place to ensure the care and integrity of items during accessioning. These should be developed from and reflect the institutional policies and procedures on security (see **Storage and Preservation**).

#### 4.2.4 Cataloguing and Documentation Procedures

##### Cataloguing

Each institution normally identifies its own minimum standard of information required for catalogued items in the collection. Each institution can also identify its preferred levels of cataloguing information and documentation for acquisitions and may notify and encourage suppliers or depositors to supply this information through the deposit or acquisition process. Staff review and revise supplied documentation to ensure it conforms to institutional guidelines and they generate catalogue records for deposited data incorporating cataloguing and documentation standards to ensure that information about those items can be made available to users through appropriate catalogues. In many cases the contextual information for resources will be crucial to their future use and this aspect of documentation should not be overlooked.

The level of cataloguing and documentation accompanying or subsequently added to an item, and any limitations these may impose can be documented for the benefit of future users. Where data resources are managed by third parties but made available via an institution, documentation may be supplied by the third party in an agreed form which conforms to institution guidelines or in the supplier's native format.

##### Retrospective documentation or catalogue enhancement

Where a need for enhanced access exists, an Institution may undertake to enhance documentation and cataloguing information to a higher standard to meet new requirements. Retrospective documentation or catalogue enhancement should also occur when the validating or audit of the documentation and cataloguing for a resource shows this to be below a minimum acceptable standard.

##### Edition and version control

Procedures for updating and edition control of any dynamic data resources accessioned (e.g. annual snapshots of databases which are regularly being updated) or for version control of accessioned items where appropriate (e.g. items accessioned in different formats or for which different formats for preservation and access had been generated.)

##### Cataloguing and documentation standards

Data documentation is essential in order effectively to exchange information and documents between platforms and individuals. At a minimum, it should provide information about an item's provenance and administrative history (including any data processing involved since its creation), contents, structure, and about the terms and conditions attached to its subsequent management and use.

It should be sufficiently detailed to support:

- Resource discovery (e.g. the location of a resource which is at least briefly described along with many other resources).

- Resource evaluation (e.g. the process by which a user determines whether s/he requires access to that resource).
- Resource ordering (e.g. that information which instructs a user about the terms and conditions attached to a resource and the processes or other means by which access to that resource may be acquired).
- Resource use (e.g. that information which may be required by a user in order to access the resource's information content).
- Resource management (e.g. administrative information essential to a resource's management and preservation as part of a broader collection and including information about location, version control, etc).

### Processing times

Ideally targets should be set and monitored for the maximum time between acquisition and cataloguing to prevent backlogs of unprocessed and potentially at risk materials developing during the accessioning process.

### Summary of recommendations

#### Transfer procedures

- Provide documentation to guide and support transfer of digital resources from suppliers.
- Decide how your transfer procedures can best be developed to support your storage and preservation policies.

#### Procedures to prepare data and documentation for storage and preservation

- Unique numbering of each item accessioned.
- Marking and labelling procedures.
- Handling guidelines for different media.
- Validation procedures to check media, content, and structure.
- Re-formatting of file or storage media formats according to agreed guidelines.
- Generating multiple copies of an item as part of an institution's storage and preservation policy.
- System and physical security policy and procedures for items during accessioning.

## Procedures for cataloguing and documentation

- A minimum standard of information required for cataloguing.
- Guidelines for retrospective documentation or catalogue enhancement.
- Procedures for updating, and managing versions or editions of an item.
- Procedures to update collection management databases.
- Selection of cataloguing and documentation standard.
- Targets for accessioning tasks and timescales for their completion.

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

### 4.3 Storage and Preservation

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

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.

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



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

Device	Operating	Non-Operating	Long term storage
<b>Magnetic tape cassettes 12.7mm</b>	18 to 24°C 45 to 55% RH	5 to 32°C 5 to 80% RH	18 to 22°C 35 to 45% RH
<b>Magnetic tape cartridges</b>	10 to 45°C 20 to 80% RH	5 to 45°C 20 to 80% RH	18 to 22°C 35 to 45% RH
<b>Magnetic tape 4 &amp; 8mm helical scan</b>	5 to 45°C 20 to 80% RH	5 to 45°C 20 to 80% RH	5 to 32°C 20 to 60% RH
<b>CD-ROM</b>	10 to 50°C 10 to 80% RH	-10 to 50°C 5 to 90% RH	18 to 22°C 35 to 45% RH

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 Reference 7**) 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 **Reference 8**) and Rothenberg (see **Reference 7**), (**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

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



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**

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

- 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**

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

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<sup>6</sup>.
- 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.

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

- 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**

#### 4.4 Metadata and Documentation

Archives, libraries and museums have always organised their collections to enable users to find the information they need more readily. This function is equally important in the digital environment where the speed of development and uptake of the Internet as a publishing medium has made the discovery of quality resources increasingly hazardous. Much work has been undertaken to develop standardised means of discovering online resources, most notably the fifteen elements represented in Dublin Core (**see Reference 31**). Increasingly, attention is being turned to the crucial role of documentation and metadata to facilitate the preservation of digital resources. Just as metadata to support resource discovery is not a new phenomenon, neither is the importance of documentation in preservation programmes:

"Documentation has always played a key role in preservation practice. This is not just a matter of academic interest: to manage collections or individual items one needs to know what one is dealing with. There are many instances where documentation provided the only information about processes that had been applied and might need to be corrected (**see Reference 32**)."

While the concept is not new, there are factors which make documentation particularly critical for the continued viability of digital materials and they relate to fundamental differences between traditional and digital resources:

- **Technology** Digital resources are dependent on hardware and software to render them intelligible. There are many potential permutations of technical requirements which need to be documented so that decisions on appropriate preservation and access strategies may be made.
- **Change** The resource cannot be preserved as a single physical entity over time. The information it contains will need to be separated from its physical carrier and moved across different technological platforms if it is to remain accessible. This will inevitably produce changes which may or may not significantly affect the integrity and/or functionality of the resource.  
Documentation of actions taken on a resource and changes occurring as a result will provide a key to future managers and users of the resource.

- **Rights management** While traditional resources may or may not be copied as part of their preservation programme, digital resources must be copied if they are to remain accessible. Managers need to know that they have the right to copy for the purposes of preservation, what (if any) devices to control rights management, such as encryption, have been used, and what (if any) implications there are for controlling access.
- **Continuity** There will be many different decision-makers and operators and quite possibly different institutions influencing the management of digital materials across time. While traditional materials may be preserved by predominantly passive preventive preservation programmes, digital materials will be subject to repeated actions over a prolonged period of time.
- **Accountability** Documentation provides an audit trail of decisions affecting the long-term viability of the material.
- **Authenticity** Documentation may be the major, if not the only, means of reliably establishing the authenticity of material following changes.
- **Cost** It will be more complex and therefore more costly to maintain access to digital materials without documentation describing its technical characteristics.
- **Feasibility** It may not be possible to recreate the material without adequate documentation or at least not cost-effective to undertake complex restoration required as a result.
- **Future** Re-use.

Additional issues needing to be resolved are:

- **Costs** Given the complexity of digital materials and their requirements for preservation, it can be assumed that only a relatively limited set of essential preservation metadata can be automatically generated. This leads to questions of to what extent there may be overlapping needs of creators/owners and those taking on responsibility for long-term preservation of the resource:
  - What metadata needs to be/can be provided by creators/owners?
  - What will need to be/can be provided by the repository accepting preservation responsibility?
  - What is the most efficient and cost-effective means of gathering all necessary metadata and documentation prior to or simultaneously with ingest/acquisition?

- What are the most efficient and effective means of ensuring that all necessary documentation and metadata is preserved along with the digital resource itself?

### What still needs to be done?

While much progress has been made in defining what is required, actually ensuring that the information is readily and cost effectively accessible remains problematic. The technical environment changes so rapidly that software can become outdated before the repository undertakes responsibility. If a third party is undertaking responsibility for preservation the issue can become even more urgent when not even corporate memory is available to help unravel the puzzle.

Two recent studies have both drawn attention to the major obstacle of locating relevant hardware, software, and format documentation.

"Documentation for hardware and software initially ubiquitous when products are first released become increasingly difficult (and in some cases impossible) to locate over time. A concerted effort should be undertaken to collect documentation, including designs." (**Ross and Gow 1999**)

An investigation undertaken by Cornell (**Lawrence 2000**) found that successful migration programs were significantly hampered by the disparity between openly published file format specifications and the increasing use of modifications to the basic standard, the latter being rarely, if ever, publicly available. Their conclusion was that:

"There is a real and pressing need to establish reliable, sustained repositories of file format specifications, documentation, implementation guides, and related software. Cornell recommends the establishment of such repositories as a prerequisite to the development of an effective national strategy." (see **Reference 35**)

Until these recommendations have been implemented, it will continue to be a hazardous and time consuming task successfully to preserve digital resources.

See **Exemplars and Further Reading**

### 4.5 Access

There has always been a strong link between preservation and access. The major objective of preserving the information content of traditional resources is so that they can remain accessible for future as well as current generations. The link is more explicit in the digital environment in that decisions on how to provide access and how to preserve a digital resource should be made, ideally, simultaneously. As well as the timing of decisions regarding preservation and access, there is also the fact that there is little point in preserving either the container or the bit stream of digital resources. To preserve access to them is also the key objective of digital preservation programmes but requires more active management throughout the lifecycle of the resource before it can be assured. While there is a strong link between preservation and access in terms of the overriding objective of a digital preservation



programme, there is also a need to make a clear distinction between them. There may be a combination of technical, legal, and pragmatic reasons to separate the access copy from the preservation copy. This section looks at some of the implications for preservation which may need to be considered when developing an access strategy.

### Storage and security

There needs to be both system and physical security if access is to be preserved over time. If the access copy is the only copy of a digital resource, then the danger of loss from theft or damage is clearly very high. In some instances, for example if large quantities of heterogeneous digital resources are being deposited with an institution, a pragmatic decision may have been made to maintain a single copy. If this approach is taken a risk assessment needs to be undertaken consisting of some of the following questions:

- Is it possible to obtain another copy of the resource from another source at any stage in the event of loss or damage? If No, make backup copy.
- Has a legal undertaking been made to preserve the resource? If Yes, make backup copy.
- Is the informational content in the resource rare or unique? If Yes, make backup copy.

See also **Acquisition and Appraisal** and **Storage and Preservation**.

### Legal

There are two main options for acquiring digital resources from external sources:

1. Via either purchase or legal requirement to deposit. This model is almost exactly analogous to the traditional model, except that additional negotiations regarding access and preservation need to be undertaken. Questions here relate to what access conditions have been permitted by the owners of the data. The ease with which digital resources can be copied and networked can be both a spur and an inhibitor to access as owners of the data seek means of ensuring unauthorised access is not permitted. If responsibility has been taken to preserve the resource and the resource is subject to copyright, it may be necessary to restrict access either for a defined period of time and/or to standalone PCs. There is much debate on the most appropriate means of adapting legal frameworks developed for traditional materials to the digital environment. This handbook is primarily concerned with encouraging thought to be given to how to manage access sanctions which might be imposed, not whether or not they should exist.
2. License digital content for an agreed period of time. This is an increasingly prevalent model and one which is well suited to the digital environment where access is not dependent on physical custody. However, there are clearly issues regarding sustained access with, to use Ann Okerson's phrase, "the possibility of uncoupling ownership from access, the material object from its intellectual content" (**Okersen 1992**). Much work has been done to try to streamline licenses, bringing obvious administrative

benefits for both publishers and institutions. Model licences, such as the one developed by NESLI for electronic journals, state that publishers must continue to provide access to material previously paid for if a subscription is cancelled. However, it is important to be aware that, even when a supplier agrees to the concept of "perpetual access" this is not completely synonymous with digital preservation, though it does at least provide greater assurance of access for the foreseeable future (see also **Preservation Issues and Rights Management**).

## Media

Depending on current and anticipated levels of use, it may be more practical to have copies stored offline, nearline, or online. What policies and procedures need to be in place to decide which of these is most appropriate, and how the resource can be preserved regardless of where it is stored?

## Technical

The large file sizes associated with uncompressed formats may make access time unacceptably slow. Similarly some formats may be more suited to presentation and therefore access but not necessarily appropriate for long-term preservation.

See **Exemplars and Further Reading**

## 4.6 Exemplars and Further Reading

There are numerous excellent sources of guidance covering all aspects of digital materials creation. This is an area where there is not only theoretical guidance but recommendations based on solid practical experience. There is now such a rich source of information relating to creating digital materials, that it is possible to avoid many of the pitfalls experienced by earlier projects. The difficulty, particularly for those new to the task, is in selecting which of the bewildering array of resources best suits a specific need. The purpose of this further reading list is to simplify the task of finding the resource most suited to a particular situation by categorising key guidance documents and supplying sufficient descriptive information to assess their relevance.

**Corporate policies and strategies - guidance**

**Corporate policies and guidelines - institutional models**

**General guidance - digital resource creation**

**Digitisation: outsourcing versus in-house**

**Guidance on selection for digitisation**

**Funding**

**Preservation metadata**

**Technical standards**

**Digitisation Costs****Training****Cataloguing and Documentation Procedures****Storage and Preservation****Metadata and Documentation****Access****Corporate policies and strategies - guidance**

1. Beagrie, N. and Greenstein, D. (1998). A Strategic Policy Framework for Creating and Preserving Digital Collections. Version 4.0 (Final Draft). ELib Supporting Study P3. Library Information Technology Centre, South Bank University, London. Also available online at: <http://ahds.ac.uk/strategic.htm>  
**Update 26-11-2004**  
Version 5 of this document is now available in either PDF or Word format at:  
<http://www.ahds.ac.uk/about/publications/index.htm>  
The study aims to provide a strategic policy framework for the creation and preservation of digital resources, and to develop guidance based on case studies, further literature and ongoing projects which will facilitate effective implementation of the policy framework. The authors advocate the concept of a life-cycle approach in preserving digital resources and suggest that the ability to preserve digital resources into the long-term rest heavily on decisions made at different stages in the life-cycle. Decisions taken during design and creation and those taken when a resource is accessioned into a collection are considered the most influential. Case studies are divided into six categories: data banks; digitisers; funding and other agencies; institutional archives; academic data archives; legal deposit libraries.
2. DLM Forum. (1997). Guidelines on Best Practice for Using Electronic Information. <http://europa.eu.int/ISPO/dlm/documents/gdlines.pdf> [PDF]  
**Update 19 March 2008**  
No longer available - information at  
<http://ec.europa.eu/archives/ISPO/dlm/>  
Designed as multidisciplinary guidelines and arising out of the DLM Forum, a European forum which brings together experts from industry, research, administration and archives to discuss issues of mutual concern. The Guidelines are intended to help define short and medium term strategies for managing electronically stored data. Annex 8.4 contains a checklist for an electronic information strategy.
3. Electronic Records Management: Framework for Information Age Government. March 2000. <http://www.e-envoy.gov.uk/assetRoot/04/00/22/94/04002294.rtf>  
Update 26-11-2004  
Electronic Records Management: Framework for Information Age Government. April 2000.

<http://e-government.cabinetoffice.gov.uk/assetRoot/04/00/22/94/04002294.rtf>

Update 12 Mar 2007

Replaced with link to PDF

[http://archive.cabinetoffice.gov.uk/e-envoy/resources-pdfs/\\$file/erm.pdf](http://archive.cabinetoffice.gov.uk/e-envoy/resources-pdfs/$file/erm.pdf)

Developed to provide guidance to departments which will need to commence working towards the Modernising Government target of all newly created public records being both electronically stored and retrieved by 2004.

4. Hedstrom, M. and Montgomery, S. (1998). Digital Preservation Needs and Requirements in RLG Member Institutions. Mountain View, CA: RLG.  
<http://www.rlg.org/preserv/digpres.html>  
Fifty-four RLG member institutions were surveyed for this study, including seven in the UK. One of the key questions the study was charged with answering was what policies and practices are being used to preserve digital materials. The report concluded that digital preservation policies are not well developed in member institutions and that "....among those institutions with digital preservation responsibilities, the lack of good models for digital preservation and confusion about the most appropriate methods and approaches are major obstacles to developing effective policies and practices."
5. Public Record Office. (1999). Guidelines on the Management, Appraisal and Preservation of Electronic Records. Volumes 1: Principles, and 2: Procedures. Kew: Public Record Office.  
<http://www.nationalarchives.gov.uk/electronicrecords/advice/guidelines.htm>  
These two publications are part of a series of guidance documents prepared under the auspices of the Electronic Records from Office Systems (EROS) programme of the Public Record Office. Volume 1, Principles, sets the scene for electronic record management and provides broad strategies arising out of the principles. Volume 2: Procedures, provides more detailed guidance for putting the principles into practice. Chapter 5 of this volume contains guidance on how to develop a preservation strategy and advises that the plan must be agreed by three parties, the systems administrator, the records manager, and the budget holder. This advice reinforces the need to a) involve what may well be administratively separate parts of the organisation in the development of an effective strategy (for example, Records and IT Managers) and b) the importance of corporate ownership of the strategy.
6. RLG/DLF Task Force on Policy and Practice for Long-Term Retention of Digital Materials.  
<http://www.rlg.org/preserv/digrlgdlf99.html>  
This Task Force was formed in response to the findings of the RLG survey (see above Hedstrom and Montgomery, 1998). The model policies and practices gathered by the Task Force relate to three categories of digital object:
  1. institutional records in digital form;
  2. locally digitised materials;
  3. electronic publications.

**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>**Corporate policies and guidelines - institutional models**

1. Columbia University Libraries. Policy for Preservation of Digital Resources. July 2000.  
<http://www.columbia.edu/cu/lweb/services/preservation/dlpolicy.html>  
Specifies different categories of digital resources the University accepts responsibility for, including "Digital Resources reformatted by CUL, and deemed to be of long-term value in digital form". Also points to other relevant policy documents, such as Selection Criteria for Digital Imaging Projects  
<http://www.columbia.edu/cu/libraries/digital/criteria.html> and Technical Recommendations for Digital Imaging Projects  
<http://www.columbia.edu/acis/dl/imagespec.html>  
**Update 24 October 2008**  
<http://www.columbia.edu/cu/libraries/digital/criteria.html> no longer valid
2. Library of Congress. Preservation Digital Reformatting Program. 1999.  
<http://lcweb.loc.gov/preserv/prd/presdig/presintro.html>  
The digitising component of the preservation reformatting program has three parts, 1) selection criteria, 2) digital reformatting principles and specifications (includes, for example, retention of analogue version of digitally reformatted items "...until the Preservation Directorate has confidence that the life-cycle management of digital data will ensure access for as long as, or longer than, the analogue version.", 3) life-cycle management of LC digital data (a term used in preference to digital preservation in order to avoid potential confusion of definition). As used here, life-cycle management is defined as "the progressive technology and workflow requirements needed to ensure long-term sustainability of and accessibility to digital objects and/or metadata."
3. National Library of Australia Digitisation Policy 2000-2004.  
<http://www.nla.gov.au/policy/digitisation.html>  
Described as "...a guide to both the digitisation of items held by the Library, and the management of these digital objects". While the policy covers four years, it includes a stated intention to review it annually. Also includes specific goals for the first year, which include setting up a Digitisation Steering Committee.
4. National Library of Canada. Networked Electronic Publications Policy and Guidelines. October 1998.  
<http://www.collectionscanada.ca/9/8/index-e.html>  
**Update 17 September 2007**

Link new location

[http://epe.lac-bac.gc.ca/100/200/301/nlc-bnc/networked\\_epubs-ef/9/8/index-e.html](http://epe.lac-bac.gc.ca/100/200/301/nlc-bnc/networked_epubs-ef/9/8/index-e.html)

5. Society of American Archivists. The Preservation of Digitized Reproductions. 1997. Online.

<http://www.archivists.org/statements/digitize.asp>

This differs from the models above in that it is intended as advice to institutions, as opposed to being tailored specifically to an individual institution. As such it tends to refer to broad principles at a fairly high level. This is probably one of the earliest explications of the role of the creator as recommended by the US Task Force. "Responsibility for long-term access to digital archives rests initially with the creator or owner of the materials. The resource and administrative implications of this fact cannot be minimized and must play a role in the decision to digitize archival and manuscript materials."

#### 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>

### General guidance - digital resource creation

1. AHDS Guides to Good Practice

<http://www.ahds.ac.uk/creating/guides/index.htm>

The Arts and Humanities Data Service has produced a series of Guides to Good Practice to provide the arts and humanities research and teaching communities with practical instruction in applying recognised standards and good practice in the creation, preservation and use of digital materials. Some of the titles are geared towards specific disciplines while others are cross-disciplinary and geared towards providing general guidance. Titles linked online at 14 June 2001:

1. Archiving Aerial Photography and Remote Sensing Data;
2. Excavation and Fieldwork Archiving;
3. GIS (Geographic Information Systems);
4. Digitising History: a guide to creating digital resources from historical documents;
5. Creating Digital Performance Resources;
6. Creating and Documenting Electronic Texts;
7. Creating Digital Resources for the Visual Arts: Standards and Good Practice;

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

2. Guides to Quality in Visual Resource Imaging. July 2000.  
<http://www.rlg.org/visguides/>  
The five guides have been commissioned by DLF and CLIR and published with RLG in order to fill a perceived gap. "While resources for instruction in digitizing text or text and images existed and were growing, none specifically addressed the challenges of two- and three-dimensional, as well as color intensive, materials. The five guides are:
  1. Planning an Imaging Project, by Linda Serenson Colet, Museum of Modern Art
  2. Selecting a Scanner, by Don Williams, Eastman Kodak Company
  3. Imaging Systems: the Range of Factors Affecting Image Quality, by Donald D'Amato, Mitretek Systems
  4. Measuring Quality of Digital Masters, by Franziska Frey, Image Permanence Institute, Rochester Institute of Technology
  5. File Formats for Digital Masters, by Franziska Frey.
3. Joint RLG and NPO Preservation Conference on Guidelines for Digital Imaging. 28-30 September 1998.  
<http://www.rlg.org/preserv/joint/>  
Gathers together a number of international experts to discuss the full range of preservation issues associated with digital imaging.
4. Kenney, Anne R. and Chapman, Stephen (1996). Digital Imaging for Libraries and Archives. New York: Cornell University Library. Ordering details online. Available: <http://www.library.cornell.edu/preservation/dila.html>  
This guide won the 1997 Leland Prize from the Society of American Archivists for "writing of superior excellence and usefulness in the field of archival history, theory and practice." The companion volume Moving Theory Into Practice (2000) is cited below.
5. Kenney, Anne R. and Rieger, Oya Y. (2000). Moving Theory Into Practice. Mountain View, CA: Research Libraries Group. (ISBN 0-9700225-0-6) Table of Contents and Ordering details online.  
<http://www.rlg.org/preserv/mtip2000.html>  
The authors have used extensive practical knowledge to provide detailed guidance to institutions contemplating digital conversion of cultural resources.
6. Higher Education Digitisation Service (HEDS).  
<http://heds.herts.ac.uk>  
This JISC funded service provides a full range of services, from advice and consultancy to actual scanning. Their website also contains links to papers prepared by HEDS staff and others.
7. Lee, S. (1999). Scoping the Future of the University of Oxford's Digital Library Collections: Final Report.

<http://www.Bodley.ox.ac.uk/scoping/report.html>

The aims of this project were:

- To document, analyse, and evaluate Oxford's current digitisation activities, as a basis for assessing the effectiveness of the various methodologies used.
- To investigate the possibilities for building on the existing project-based work and for migrating it into viable services for library users.
- To develop appropriate selection criteria for creating digital collections in the context of local, national, and international scholarly requirements for digital library products and services.
- To make recommendations for further investment and activity within the UK research libraries community.

The resulting report, with ten appendices, is an extremely detailed investigation of a whole range of issues amounting to a strategic plan for the future digital library development of this institution. While specifically designed for one university, the issues are also applicable to many other organisational contexts.

8. NOF-digitise Technical Standards and Guidelines. Version One; June 2000.

<http://www.peoplesnetwork.gov.uk/content/technical.asp>

**Update 27-01-2006**

This link no longer active. Version One and later versions available at:

[http://www.mla.gov.uk/webdav/harmonise?Page/@id=73&Document/@id=18612&Section\[@stateId\\_eq\\_left\\_hand\\_root\]/@id=4332](http://www.mla.gov.uk/webdav/harmonise?Page/@id=73&Document/@id=18612&Section[@stateId_eq_left_hand_root]/@id=4332)

Stage two of the nof-digitise programme has provided £50 million worth of funding for the creation of digital content for users of the People's network and the national grid for learning. A web-based resource has been provided to support Stage two, providing advice on standards for accessibility, availability, document and file formats, search and request protocols, security and e-commerce, preservation and metadata.

9. Technical Advisory Service for Images (TASI).

<http://www.tasi.ac.uk>

A JISC funded service set up to advise and support the academic community on the digital creation, storage and delivery of image-related information. TASI also collaborated with the Visual Arts Data Service (VADS) to produce Creating Digital Resources for the Visual Arts, one of the titles in the AHDS Guides to Good Practice series.

**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>

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



### Digitisation: outsourcing versus in-house

Decisions will need to be made on whether to outsource all or part of a digitisation project or to undertake all of it in-house. Some guidance can be found in Kenney and Chapman (1996), Chapter 5, Outsourcing Imaging Services which reviews the pros and cons of outsourcing. Tanner and Lomax-Smith (1999) suggest that while cost is likely to favour outsourcing if large volumes of material are being digitised, other factors, such as whether or not the material can be taken out of the institution, will obviously influence whether in-house digitisation will be the preferred option.

Whatever option is selected, the host institution will still need to commit significant resources to ensuring the project successfully delivers its stated goals.

1. Kenney, Anne R. and Chapman, Stephen (1996). Digital Imaging for Libraries and Archives. New York: Cornell University Library. Ordering details online.  
<http://www.library.cornell.edu/preservation/dila.html>
2. Tanner, S. and Lomax-Smith, J. (1999). 'How Much Does It Really Cost?' Paper for DRH '99 Conference.  
<http://heds.herts.ac.uk/resources/papers/drh99.pdf> [PDF]  
**Update 09 Nov 2007**  
Paper no longer available online
3. RLG Tools for Digital Imaging  
<http://www.rlg.org/preserv/RLGtools.html>

RLG Tools for Digital Imaging provide guidance in the form of:

- A worksheet for Estimating Digital Reformatting Costs.
- RLG Guidelines for Creating a Request for Proposal.
- RLG Model RFI (an example of how Cornell University invited vendors that would be interested in receiving their project RFP).
- RLG Model RFP (an example of how Cornell University adapted the RLG Guidelines for Creating a Request for Proposal for use in a text-based digitisation project).

#### 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>

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

### Guidance on selection for digitisation

1. Ayris, Paul. (1998). 'Guidance for selecting materials for digitisation'. Joint RLG and NPO Preservation Conference: Guidelines for Digital Imaging.  
<http://www.rlg.org/preserv/joint/ayris.html>  
The paper identifies studies which have considered the role of selection in the process of digitisation and suggests a decision-making matrix of twenty questions grouped around four issues, Assessment; Gains; Standards; Administrative Issues.
2. Lee, S. (1999). Scoping the Future of the University of Oxford's Digital Library Collections: Final Report.  
<http://www.Bodley.ox.ac.uk/scoping/report.html>  
Appendix B is a detailed workflow chart providing assistance in deciding whether or not to proceed with a digitisation project.
3. Hazen, D., Horrell, J. and Merrill-Oldham, J. (1998). Selecting Research Collections for Digitization. Council on Library and Information Resources, August 1998.  
<http://www.clir.org/pubs/reports/reports.html>

Provides detailed planning information for research libraries contemplating large-scale digital conversion of holdings for research and teaching purposes. Discusses selection criteria, imaging standards, rights management issues, preservation concerns, and impact of digitisation on the library and its users.

#### 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>

### Funding

1. The Technical Advisory Service for Images (TASI) maintains a list of potential sources of funding on its website.  
<http://www.tasi.ac.uk/resources/funding.html>

Note that some of these have specific clauses relating to digital preservation. For example the Arts and Humanities Research Board (AHRB) makes it a condition that for grants awarded where a significant product or by-product is the creation of an electronic resource, it is offered for deposit at the Arts and Humanities Data Service (AHDS) within three months of the end of the project. Both time and adequate funding are provided to prepare the data for deposit (see AHRB Grant Applications and Awards: <http://www.ahds.ac.uk/ahrb/>). The New Opportunities Fund (NOF) advises in their Information for Applicants that intellectual property issues and technical standards

identified by NOF must be observed (see New Opportunities Fund, information for new applicants at <http://www.nof-digitise.org> **Update 20-09-2005** This link no longer active. The NOF has now been replaced by the Big Lottery Fund: <http://www.biglotteryfund.org.uk/>).

The deadlines for responding to calls for proposals may not always mean that it is feasible to include all costs for the project, and in particular for keeping the data but it does need to be acknowledged that this will become a cost to the institution. See also **Costs**.

**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>

**Preservation metadata**

1. National Archives of Australia. Recordkeeping Metadata Standard for Commonwealth Agencies. May 1999.  
<http://www.naa.gov.au/recordkeeping/control/rkms/summary.htm>  
**Update 17 September 2007**  
This document no longer available online
2. Bearman, David and Sochats, Ken. (1996). Metadata Requirements for Evidence. Pittsburgh, Pa: University of Pittsburgh School of Information Science.  
<http://www.archimuse.com/papers/nhprc/BACartic.html>
3. Cedars Project Team and UKOLN. Metadata for Digital Preservation: the Cedars Project Outline Specification. Draft for Public Consultation. March 2000  
<http://www.leeds.ac.uk/cedars>  
This document represents a major aspect of the work of Cedars in the development of a metadata framework which will enable the long-term preservation of digital resources. The outline indicates that it generally adheres to the metadata identified by the Reference Model for an Open Archival Information System (OAIS). The document "starts with the structure provided by the OAIS model and populates it with metadata elements chosen by practical investigation of archiving real digital resources, further refined by comments received from a selective consultation process." It also restricts itself to metadata required for preservation, rather than other processes.
4. Dollar, Charles. (1999). Authentic Electronic Records: Strategies for Long-Term Access. Chicago: Cohasset Associates. (ISBN 0-9700640-0-4) Appendix 7. Preservation Metadata Model.

5. National Library of Australia. Draft Preservation Metadata Set. October 1999  
<http://www.nla.gov.au/preserve/pmeta.html>  
This has been developed as part of the NLA's plans for its digital collections. The introduction states that " There have been a number of efforts to develop metadata specifications and sets to support a wide variety of digital resources. Because of its pressing business needs to manage both 'born digital' and 'digital surrogate' collections, the National Library of Australia has tried to find, or if necessary develop, metadata models to accommodate both." The draft also emphasises that the metadata set is intended as a data output model, i.e. information required to manage digital collections, not necessarily what data should be entered, how it should be entered, by whom, and at what time. Like the Cedars specification, this document restricts itself to metadata required for preservation.
6. Public Record Office. (1999). Guidelines on the Management, Appraisal, and Preservation of Electronic Records. Volume 2: Principles. Chapter 2.  
<http://www.nationalarchives.gov.uk/electronicrecords/advice/guidelines.htm>  
Defines three classes of metadata: document metadata; record level metadata; and file/folder metadata and recommends elements for each.
7. RLG Working Group on Preservation Issues of Metadata; Final Report. May 1998.  
<http://www.rlg.org/preserv/presmeta.html>  
The Working Group noted that to date, the emphasis of metadata has been on resource discovery and retrieval. Taking two prominent metadata systems, Dublin Core and the Program for Cooperative Cataloguing's USMARC-based core record standard, the group specified those elements not covered by these two systems but important to serve the preservation needs of digital masters. The group confined itself to digital image files and recommended sixteen data elements for this category of digital resource.

See also **Metadata and Documentation**.

#### 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>

### Technical standards

1. Beagrie, N. and Greenstein, D. (1998). Managing Digital Collections: AHDS Policies, Standards and Practices. Consultation Draft. December 1998.  
<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> [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 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. Public Record Office (Victoria). Standard for the Management of Electronic Records. PROS 99/007. Version 1.0 April 2000.

<http://www.prov.vic.gov.au/vers/standards/pros9907/99-7toc.htm>

**Update 26-11-2004**

Version 2 now available: Management of Electronic Records PROS 99/007 (Version 2)

<http://www.prov.vic.gov.au/vers/standard/version2.htm>

Designed for the Victorian public sector records (but with much that is applicable at a more global level) the standard provides three specifications which provide:

- 1) technical detail about the long-term preservation of electronic records;
- 2) the requirements for records management systems which maintain electronic records; and
- 3) the metadata required for the proper management and retention of electronic records.

4. TASI. Advice: Creating Digital Images.

<http://www.tasi.ac.uk/advice/advice.html>

Includes general advice on selecting file formats for images.

#### 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>

## Digitisation Costs

### Specific Case Studies of Digitisation Projects

1. Internet Library of Early Journals (ILEJ). (1999). <http://www.bodley.ox.ac.uk/ilej>

This eLib project involved the universities of Birmingham, Leeds, Manchester, and Oxford and investigated the feasibility of digitising substantial runs of selected 18th and 19th century British journals. The final report of the project makes a number of observations and recommendations. The chapter on costs indicates an estimated cost of £4.21 per indexed page image accessible on the Internet but notes that "this estimate of expenditure does not take into account the contribution of the IT and library infrastructure of the four institutions". Archiving costs have been estimated at £20 per Gb per annum, totalling around £2,400 p.a. for the ILEJ project. The long-term future and funding of this was still unclear in June 2000.

2. Lee, S. (1999). Scoping the Future of the University of Oxford's Digital Library Collections: Final Report.  
<http://www.Bodley.ox.ac.uk/scoping/report.html>

Appendix E, Digitization Method, includes examples of costs established from projects, including JIDI, BUILDER, and the Wilfred Owen project. It also cites examples of the cost-effectiveness of outsourcing.

### General guidance on estimating costs of creating digital surrogates

3. Puglia, S. (1999). 'The Costs of Digital Imaging Projects'. RLG DigiNews 3(5) October 15 1999.  
<http://www.rlg.org/preserv/diginews/diginews3-5.html>

Averages data from a number of individual projects and estimates that for production costs, approximately one-third are for digital conversion, slightly less than one-third are for metadata creation, slightly more than one-third are for other activities, such as administration and quality control. This article also makes the important point that long-term maintenance of the digital images and associated metadata is often not considered as part of project costs but needs to be planned for. However, the article also notes that there are few models for estimating these and they vary considerably.

4. Tanner, S. and Lomax-Smith, J. (1999). 'How Much Does It Really Cost? Paper for DRH '99 Conference.  
<http://heds.herts.ac.uk/resources/papers/drh99.pdf> [PDF]  
**Update 09 Nov 2007**  
Paper no longer available online

This article provides general advice on project management and includes a matrix of potential cost factors which can be used to estimate the costs of a digitisation project.

### Comparative Costs of Digitisation, Microform and Paper

5. Kingma, B. (1999). The Economics of Digital Access: the Early Canadiana Online Project. <http://www.canadiana.org/eco/english/kingma.pdf>

**Update 26-11-2004**

A 2000 version of "The Costs of Print, Fiche, and Digital Access The Early Canadiana Online Project" by Bruce R. Kingma can be found at

<http://www.dlib.org/dlib/february00/kingma/02kingma.html>

This is an extremely detailed but highly specific investigation into the comparative costs of digital, print and microfiche access for the early Canadiana online project. The purpose of the project was "to lay the groundwork for the costing and creation of a Canadian digital collection and database to be made available on the Internet." This report effectively amounts to a detailed business case for making rare collections available via the Internet. It concludes that the cost of digital information is lower on a cost per library or per patron basis so long as a sufficient number of libraries are interested in subscribing to the database.

**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>

### Training

#### Creating/managing digital resources

The following organisations offer access to a range of training courses and workshops relating to creating and managing digital resources either organised themselves and/or linking to others.

1. Arts and Humanities Data Service. Website: <http://ahds.ac.uk>
2. Higher Education Digitisation Service (HEDS) Website: <http://heds.herts.ac.uk>
3. Humanities Advanced Technology and Information Institute (HATII), University of Glasgow. Website: <http://www.hatii.arts.gla.ac.uk>
4. Humanities Computing Unit, University of Oxford. Website: <http://www.hcu.ox.ac.uk>  
**Update 21-12-2004**  
Link to the website disabled as HCU no longer in existence.

5. Technical Advisory Service (TASI) Website:  
<http://www.tasi.ac.uk>

**Search Other Resources**

Search Archives of Digital Preservation Jiscmail list

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

Search Preserving Access to Digital Information (PADI) Gateway

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

**Electronic records management**

6. Public Record Office  
Website:  
<http://www.nationalarchives.gov.uk/recordsmanagement/training/>  
Training and consultancy services.

**Cataloguing and Documentation Procedures**

1. Reference Model for an Open Archival Information System (OAIS) Draft  
Recommendation for Space Data System Standard. May 1999.

**Update 21-12-2002**

This document is now available as either a pdf or word document from:

[http://ssdoo.gsfc.nasa.gov/nost/isoas/ref\\_model.html](http://ssdoo.gsfc.nasa.gov/nost/isoas/ref_model.html)

The OAIS Reference Model provides a useful overview of the "ingest" process and a high-level process model relevant to most institutions.

2. Beagrie, N and Greenstein, D. (1998). Managing Digital Collections: AHDS Policies, Standards and Practices. Consultation Draft. December 1998.

<http://www.ahds.ac.uk/about/reports-and-policies/index.htm>

Provides a summary of preferred formats recommended by AHDS service providers, and accessioning and transfer procedures use by the AHDS. It also includes two detailed case studies of accessioning in the History Data Service and the Oxford Text Archive.

3. National Archive of New Zealand. (1998). Appraisal Standard, Standard for the Appraisal of Public Records and Archives. Wellington, 1998.  
[http://www.archives.govt.nz/statutory\\_regulatory/reviews/appraisal\\_service/options\\_per.html](http://www.archives.govt.nz/statutory_regulatory/reviews/appraisal_service/options_per.html)

**Update 26-11-2004**

PDF now available

<http://www.archives.govt.nz/continuum/dls/pdfs/s1-standard-appraisal.pdf> [PDF]

**Update 21-12-2006**

PDF now available



<http://www.archives.govt.nz/continuum/documents/publications/s1/S1-appraisal-standard-introduction.shtml>

4. Public Record Office (UK). (1999). Electronic Records from Office Systems Project (EROS). Series of guides on management, appraisal and preservation of electronic records in government. Kew, Surrey, 1999.

<http://www.pro.gov.uk/recordsmanagement/eros/default.htm>

**Update 12-01-2005**

The EROS project has now ended. This redirected page contains links to relevant guidance:

<http://www.nationalarchives.gov.uk/recordsmanagement/>

**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>

## Storage and Preservation

### General overviews and guidance

1. Arms, C. (2000). 'Keeping Memory Alive: Practices for Preserving Digital Content at the National Digital Library Program of the Library of Congress'. RLG DigiNews: Volume 4 (3). June 15 2000.  
<http://www.rlg.org/preserv/diginews/diginews4-3.html>
2. Dollar, C. (1999). Authentic Electronic Records: Strategies for Long-Term Access. Chicago: Cohasset Associates. (ISBN 0-9700640-0-4).
3. AHDS: Guides to Good Practice  
<http://www.ahds.ac.uk/creating/guides/index.htm>

Titles linked in June 2001:

Archiving Aerial Photography and Remote Sensing Data; Excavation and Fieldwork Archiving; GIS (Geographic Information Systems);

Digitising History: a Guide to Creating Digital Resources from Historical Documents; Creating Digital Performance Resources; Creating and Documenting Electronic Texts; Creating Digital Resources for the Visual Arts: Standards and Good Practice.

**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>**Storage and maintenance - models and guidance**

1. 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/>
2. A Digital Preservation Strategy for the PRO. November 1999.
3. National Library of Australia. First Steps in Preserving Digital Publications. 1999.  
<http://www.nla.gov.au/pres/epupam.html>
4. 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>
5. National Library of Canada. Networked Electronic Publications: Policies and Guidelines. October 1998.  
<http://www.collectionscanada.ca/9/8/index-e.html>  
**Update 17 September 2007**  
Link new location  
[http://epe.lac-bac.gc.ca/100/200/301/nlc-bnc/networked\\_epubs-ef/9/8/index-e.html](http://epe.lac-bac.gc.ca/100/200/301/nlc-bnc/networked_epubs-ef/9/8/index-e.html)
6. NOF-digitise Technical Standards and Guidelines. Version One; June 2000.  
<http://www.peoplesnetwork.gov.uk/content/technical.asp>  
**Update 27-01-2006**  
This link no longer active. Version One and later versions available at:  
[http://www.mla.gov.uk/webdav/harmonise?Page/@id=73&Document/@id=18612&Section\[@stateId\\_eq\\_left\\_hand\\_root\]/@id=433](http://www.mla.gov.uk/webdav/harmonise?Page/@id=73&Document/@id=18612&Section[@stateId_eq_left_hand_root]/@id=433)
7. Oxford University. Policy on Computer Archiving Service. 1997.  
<http://www.oucs.ox.ac.uk/hfs/policy/>
8. Oxford University Computing Services. Preservation of the Electronic Assets of a University. 1997.  
<http://www.lmcp.jussieu.fr/eunis/html3/congres/EUNIS97/papers/052202.html>
9. PADI. 'Storage'.  
<http://www.nla.gov.au/padi/topics/10.html>

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

10. TASI. Recommendations for Digital Preservation and Storage.  
<http://www.tasi.ac.uk/delivering/digpres.html>
11. Van Bogart, J. (1995). Magnetic Tape Storage and Handling. Council on Library and Information Resources. (ISBN 1-887334-40-8).  
<http://www.clir.org/pubs/reports/pub54>

**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>**Preservation strategies****Overviews and general guidance**

1. Bearman, D. (1999). 'Reality and Chimeras in the Preservation of Electronic Records'. D-Lib Magazine. April 1999.  
<http://www.dlib.org/dlib/april99/bearman/04bearman.html>
2. Berthon, H. and Webb, C. (2000). 'The Moving Frontier: Archiving, Preservation and Tomorrow's Digital Heritage.' Paper presented at VALA 2000 - 10th VALA Biennial Conference and Exhibition, Melbourne, Victoria, 16-18 February 2000.  
<http://www.nla.gov.au/nla/staffpaper/hberthon2.html>
3. Hendley, T. (1998). Comparison of Methods & Costs of Digital Preservation. British Library Research and Innovation Report 106. London: The British Library. (ISBN 0 7123 9713 2)  
<http://www.ukoln.ac.uk/services/elib/papers/tavistock/hendley/hendley.html>
4. PADI. 'Digital Preservation Strategies'.  
<http://www.nla.gov.au/padi/topics/18.html>

**Migration**

5. Lawrence, G.W. et al. (2000). Risk Management of Digital Information: a File Format Investigation. Council on Library and Information Resources. June 2000. (ISBN 1-887334-78-5).  
<http://www.clir.org/pubs/abstract/pub93abst.html>

**Emulation**

6. CAMiLEON (Creative Archiving at Michigan and Leeds; Emulating the Old and the New) Project. Three-year NSF/JISC funded project commenced 1 October 1999. Further

details online. Available from the Cedars website:

<http://www.leeds.ac.uk/cedars>

7. Rothenberg, J. (2000). An Experiment in Using Emulation to Preserve Digital Publications. A report commissioned by the Koninklijke Bibliotheek (KB). <http://www.kb.nl/coop/nedlib/results/emulationpreservationreport.pdf> [PDF]  
**Update 11 Aug 2006**  
New location  
<http://nedlib.kb.nl/results/emulationpreservationreport.pdf>
8. Rothenberg, J. (1999). Avoiding Technological Quicksand: Finding a Viable Technical Foundation for Digital Preservation. Council on Library and Information Resources. January 1999. (ISBN 1-887334-63-7).  
<http://www.clir.org/pubs/abstract/pub77.html>

### Digital archaeology

9. Ross, S. and Gow, A. (1999). Digital Archaeology: Rescuing Neglected and Damaged Data Resources. British Library Research and Innovation Report 108. London, British Library, 1999.  
<http://www.hatii.arts.gla.ac.uk/research/BrLibrary/rosgowrt.pdf> [PDF]

### Encapsulation

10. Heminger, A. R. and Robertson, S. B. (1998). 'Digital Rosetta Stone: A Conceptual Model for Maintaining Long-Term Access to Digital Documents.' ERCIM Workshop Proceedings No. 98-W003.  
<http://www.ercim.org/publication/ws-proceedings/DELOS6/rosetta.pdf> [PDF]
11. Reference Model for an Open Archival Information System (OAIS) Draft Recommendation for Space Data System Standards, of the Consultative Committee for Space Data Systems (CCSDS), CCSDS 650.0-R-1, May 1999.  
<http://www.ccsds.org/documents/650x0b1.pdf>  
**Update 27-01-2006**  
The draft OAIS recommendation is no longer available, the full specification (2002) can be found at: <http://public.ccsds.org/publications/archive/650x0b1.pdf>

#### 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>

## Metadata and Documentation

### Documentation - standards and guidance

1. Data Documentation Initiative (DDI)  
<http://www.icpsr.umich.edu/DDI/codebook>

This is an example of an initiative by a particular community, the social science research community, to "establish an international criterion and methodology for the content, presentation, transport and preservation of metadata about data sets in the social and behavioral sciences." Social science research has for many years pioneered the re-use of data emanating from research projects.

2. The Data Archive, University of Essex. Guide to Depositing Data. Guidelines for Documenting Data.  
<http://www.data-archive.ac.uk>
3. Arts and Humanities Data Service.  
<http://www.ahds.ac.uk>

Individual service providers offer guidance on documentation, for example, the History Data Service: Guidelines for Documenting Data <http://hds.essex.ac.uk/docguide.asp> and the Archaeology Data Service, Guidelines for Depositors Version 1.1 <http://ads.ahds.ac.uk/project/userinfo/deposit.html>.

**Update 11th April 2008**

URL no longer available. Version 1.3 available at  
<http://ads.ahds.ac.uk/project/userinfo/deposit.cfm>

In addition, the Guides to Good Practice series also offer advice as well as guidance on why documentation is important. For example Creating Digital Resources for the Visual Arts: Standards and Good Practice. Section 4. Standards for Data Documentation [http://vads.ahds.ac.uk/guides/creating\\_guide/sect41.html](http://vads.ahds.ac.uk/guides/creating_guide/sect41.html) and Creating and Documenting Electronic Texts. Chapter 6: Documentation and Metadata. <http://ota.ahds.ac.uk/documents/creating/chap6.html>.

**Update 09 May 2008**

New location  
<http://ota.oucs.ox.ac.uk/documents/creating/cdet/chap6.html>

### Metadata - developing standards

4. RLG Working Group on Preservation Issues of Metadata. Final Report. May 1998.  
<http://www.rlg.org/preserv/presmeta.html>

The Working Group noted that to date the emphasis of metadata has been on resource discovery and retrieval. Taking two prominent metadata systems, Dublin Core and the Program for Cooperative Cataloguing's USMARC-based core record standard, the

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

group specified those elements not covered by these two systems but important to serve the preservation needs of digital masters. The group confined itself to digital image files and recommended sixteen data elements for this category of digital resource.

5. Reference Model for an Open Archival Information System (OAIS) Draft Recommendation for Space Data System Standard. May 1999.

**Update 21-12-2002**

This document is now available as either a pdf or word document from:

[http://ssdoo.gsfc.nasa.gov/nost/isoas/ref\\_model.html](http://ssdoo.gsfc.nasa.gov/nost/isoas/ref_model.html)

This model aims to develop a common framework for all archives, digital and non digital. However of particular relevance and interest to the understanding of digital resources is the OAIS definition of Archival Information Packages (AIP's).

This recognises and identifies the range of elements required before a digital resource is useable and reinforces the fundamental differences between preserving digital and traditional resources. An AIP consists of both content information (both the data object and any Representation Information (RI) needed to render it intelligible) and Preservation Description Information (PDI), descriptive metadata which allows the essence of what the content information is to be understood indefinitely.

6. NEDLIB (Networked European Deposit Library)

<http://www.kb.nl/nedlib>

This project has twelve partners consisting of deposit libraries, archives, and IT developers. Three publishers are also contributing to the project, which runs from January 1998 to December 2000. The main focus of NEDLIB has been on the technical aspects of digital preservation. NEDLIB has based its Deposit System for Electronic Publications (DSEP) on the OAIS model but has added a specific preservation module specifically to identify where "transformation processes" (i.e. migrations) take place.

7. Cedars Project Team and UKOLN. Metadata for Digital Preservation: the Cedars Project Outline Specification. Draft for Public Consultation. March 2000.

<http://www.leeds.ac.uk/cedars>

This document represents a major aspect of the work of Cedars in the development of a metadata framework which will enable the long-term preservation of digital resources. The outline indicates that it generally adheres to the metadata identified by the Reference Model for an Open Archival Information System (OAIS). The document "starts with the structure provided by the OAIS model and populates it with metadata elements chosen by practical investigation of archiving real digital resources, further refined by comments received from a selective consultation process." It also restricts itself to metadata required for preservation, rather than other processes.

8. National Library of Australia. Draft Preservation Metadata Set. October 1999.

<http://www.nla.gov.au/preserve/pmeta.html>

This has been developed as part of the NLA's plans for its digital collections. The introduction states that "There have been a number of efforts to develop metadata specifications and sets to support a wide variety of digital resources. Because of its pressing business needs to manage both 'born digital' and 'digital surrogate' collections, the National Library of Australia has tried to find, or if necessary develop, metadata models to accommodate both." The draft also emphasises that the metadata set is intended as a data output model, i.e. information required to manage digital collections, not necessarily what data should be entered, how it should be entered, by whom, and at what time. Like the Cedars specification, this document restricts itself to metadata required for preservation.

#### 9. RLG/OCLC

<http://www.rlg.org/pr/pr2000-oclc.html>

On March 10 2000, RLG and OCLC agreed to combine forces to work towards creating infrastructures for digital archiving. The first steps towards this wider aim are collaboration on two working documents, one on characteristics of reliable archiving services and another on preservation metadata. The draft documents will be made available on both the RLG <http://www.rlg.org/> and OCLC <http://www.oclc.org> and comments will be invited before final publication.

#### **Update 03 October 2007**

RLG has merged with OCLC <http://www.oclc.org/>

The above examples of work being undertaken in this area have all been based on practical experience and identified needs and show considerable progress is being made. Ongoing development is still needed, particularly for "published" digital resources. This is because a) it is impossible to predict precisely what will be required for heterogeneous digital resources, and b) as the above examples demonstrates, it is difficult to establish a standard set of elements satisfying the requirements of all institutions for all digital resources.

#### **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>

### **Record keeping metadata**

10. National Archives of Australia. Recordkeeping Metadata Standard for Commonwealth Agencies. May 1999.

<http://www.naa.gov.au/recordkeeping/control/rkms/summary.htm>

#### **Update 26 September 2007**

This document no longer available online

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

11. Bearman, David and Sochats, Ken. (1996). Metadata Requirements for Evidence. Pittsburgh, Pa: University of Pittsburgh School of Information Science.  
<http://www.archimuse.com/papers/nhprc/BACartic.html>
12. Dollar, Charles. (1999). Authentic Electronic Records: Strategies for Long-Term Access. Chicago: Cohasset Associates. (ISBN 0-9700640-0-4).  
Appendix 7. Preservation Metadata Model.

**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>**Access**

Many of the references cited in further reading sections elsewhere in the handbook, in particular those in **Creating Digital Materials** also include sections relevant to access. In addition, the following are references for model licences:

1. ECUP (European Copyright User Platform) Licensing Issues.  
<http://www.eblida.org/ecup/licensing/lic.htm>  
**Update 07 Mar 2007**  
Link removed.
2. Licensingmodels.com.  
<http://www.licensingmodels.com>
3. NESLI (National Electronic Site Licensing Initiative).  
<http://www.nesli2.ac.uk>

**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>**4. References**

1. Waters, D. and Garrett, J. (1996). Preserving Digital Information: Report of the Task Force on Archiving of Digital Information commissioned by the Commission on Preservation and Access and the Research Libraries Group. Washington, DC: Commission on Preservation and Access. p.47.  
<http://www.rlg.org/ArchTF/tfadi.index.htm>

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



2. DLM Forum. (1997). Guidelines on the Best Practices for Using Electronic Information. These make a useful distinction between 'live' and 'frozen' information, where the former is, in general, managed by the creator and the latter by archivists and librarians who need to ensure it remains the same.  
<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/>
3. Source of checklist: Adapted from Tanner, S. and Lomax-Smith, J. (1999). 'How Much Does It Really Cost?' Paper for DRH '99 Conference.  
<http://heds.herts.ac.uk/resources/papers/drh99.pdf>  
**Update 09 Nov 2007**  
Paper no longer available online  
and AHDS 'Digitisation: a Project Planning Checklist'  
<http://www.ahds.ac.uk/creating/information-papers/checklist/index.htm>
4. Public Record Office. (1999). Guidelines on the Management, Appraisal and Preservation of Electronic Records. Volume 2: Procedures, p. 17.  
<http://www.pro.gov.uk/recordsmanagement/eros/guidelines>  
**Update 09 Aug 2006**  
The EROS project has now ended. This redirected page contains links to relevant guidance:  
<http://www.nationalarchives.gov.uk/recordsmanagement/>
5. Kenney, A. R. and Chapman, S. (1996). Digital Imaging for Libraries and Archives. New York: Cornell University Library. Figure 1. p.190.
6. A Digital Preservation Strategy for the PRO. 1999. 3.5.3.
7. Modernising Government. March 1999. Cm 4310. Chapter 5.  
<http://www.archive.official-documents.co.uk/document/cm43/4310/4310.htm>
8. E-government: a strategic framework for public services in the Information Age.  
[http://www.e-envoy.gov.uk/EStrategy/StrategicFrameworkArticle/fs/en?CONTENT\\_ID=4000067&chk=%2BE%2BDg%2B](http://www.e-envoy.gov.uk/EStrategy/StrategicFrameworkArticle/fs/en?CONTENT_ID=4000067&chk=%2BE%2BDg%2B)  
**Update 11 May 2007**  
Link disabled – obsolete
9. The FOI bill had been given royal assent in July 2000 and a timeframe of everything being in place within five years of royal assent has been given. A Model Action Plan for Developing Records Management Compliant with the Lord Chancellor's Code of Practice under section 46 of the Freedom of Information Act 2000 is available from the PRO website at  
<http://www.pro.gov.uk/recordsmanagement/access/foiactionplan.pdf>  
**Update 10 Aug 2006**  
<http://www.nationalarchives.gov.uk/recordsmanagement/code/>

10. Public Record Office. (1999). Guidelines on the Management, Appraisal and Preservation of Electronic Records. Volumes 1: Principles, p. 17 Kew: Public Record Office.  
<http://www.pro.gov.uk/recordsmanagement/eros/guidelines>  
**Update 09 Aug 2006**  
The EROS project has now ended. This redirected page contains links to relevant guidance:  
<http://www.nationalarchives.gov.uk/recordsmanagement/>
11. Public Record Office. (1999). Guidelines on the Management, Appraisal and Preservation of Electronic Records. Volume 1: Principles, p. 25. Kew: Public Record Office.  
<http://www.pro.gov.uk/recordsmanagement/eros/guidelines>  
**Update 09 Aug 2006**  
The EROS project has now ended. This redirected page contains links to relevant guidance:  
<http://www.nationalarchives.gov.uk/recordsmanagement/>
12. DLM Forum. (1997). Guidelines on the Best Practices for Using Electronic Information. p.20.  
<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/>
13. Electronic Records Management: Framework for Information Age Government March 2000.  
<http://www.e-envoy.gov.uk/assetRoot/04/00/22/94/04002294.rtf>  
**Update 26 Nov 2004**  
Electronic Records Management: Framework for Information Age Government. April 2000.  
<http://e-government.cabinetoffice.gov.uk/assetRoot/04/00/22/94/04002294.rtf>  
**Update 12 Mar 2007**  
Replaced with link to PDF  
[http://archive.cabinetoffice.gov.uk/e-envoy/resources-pdfs/\\$file/erm.pdf](http://archive.cabinetoffice.gov.uk/e-envoy/resources-pdfs/$file/erm.pdf)
14. See Ross, S. and Gow, A. (1999). Digital Archaeology: Rescuing Neglected and Damaged Data Resources. p. 29.  
<http://www.hatii.arts.gla.ac.uk/research/BrLibrary/rosgowrt.pdf>
15. Berkeley Digital Library Sunsite.  
<http://sunsite.berkeley.edu/Admin/collection.html>
16. Beagrie N. and Greenstein D. (1998). Managing Digital Collections: AHDS Policies, Standards and Practices. Consultation draft. Version 1. 15 December 1998. 2.3.2.  
<http://www.ahds.ac.uk/about/reports-and-policies/index.htm>

17. National Library of Canada. (1998). Networked Electronic Publications: Policy and Guidelines. p. 7-8. Online. Available:  
<http://www.collectionscanada.ca/9/8/index-e.html>  
**Update 17 September 2007**  
Link new location  
[http://epe.lac-bac.gc.ca/100/200/301/nlc-bnc/networked\\_epubs-ef/9/8/index-e.html](http://epe.lac-bac.gc.ca/100/200/301/nlc-bnc/networked_epubs-ef/9/8/index-e.html)
18. Ross, S and Gow,A. (1999). Digital Archaeology: Rescuing Neglected and Damaged Data Resources. p. 3.  
<http://www.hatii.arts.gla.ac.uk/research/BrLibrary/rosgowrt.pdf>
19. van Bogart, J. (1995). Magnetic Tape Storage and Handling. Council on Library and Information Resources. (ISBN 1-887334-40-8)  
<http://www.clir.org/pubs/reports/pub54>
20. British Standards Institute. (1999). Information Security Management (BS7799-2: 1999). Order details online: <http://www.bsi-global.com/index.xalter>  
**Update 11 May 2007**  
Link disabled – obsolete
21. Data Protection Act 1998.  
<http://www.opsi.gov.uk/acts/acts1998/19980029.htm>
22. For example Arms, C. (2000). 'Keeping Memory Alive: Practices for Preserving Digital Content at the National Digital Library Program of the Library of Congress.' RLG DigiNews. Volume 4 (3). June 15 2000. p. 5.  
<http://www.rlg.ac.uk/preserv/diginews/diginews4-3.html>  
**Update 07 July 2006**  
<http://www.rlg.org/preserv/diginews/diginews4-3.htm>
23. Lawrence, G.W. et al. (2000). Risk Management of Digital Information: a File Format Investigation. Council on Library and Information Resources. June 2000. (ISBN 1-887334-78-5).  
<http://www.clir.org/pubs/abstract/pub93abst.html>
24. Rothenberg, J. (1999). Avoiding Technological Quicksand: Finding a Viable Technical Foundation for Digital Preservation. Council on Library and Information Resources. January 1999. (ISBN 1-887334-63-7).  
<http://www.clir.org/pubs/abstract/pub77.html>
25. CAMiLEON (Creative Archiving at Michigan and Leeds; Emulating the Old and the New) Project. Available from the Cedars website:  
<http://www.leeds.ac.uk/cedars/>
26. Rothenberg, J. (2000). An Experiment in Using Emulation to Preserve Digital Publications. A report commissioned by the Koninklijke Bibliotheek (KB).  
<http://www.kb.nl/coop/nedlib/results/emulationpreservationreport.pdf>

**Update 11 Aug 2006**

New location <http://nedlib.kb.nl/results/emulationpreservationreport.pdf>

27. Bearman, D. (1999). 'Reality and Chimeras in the Preservation of Electronic Records'. D-Lib Magazine. April 1999.  
<http://www.dlib.org/dlib/april99/bearman/04bearman.html>
28. 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. p. 44.  
<http://www.hatii.arts.gla.ac.uk/research/BrLibrary/rosgowrt.pdf>
29. Lawrence, G.W. et al. (2000). Risk Management of Digital Information: a File Format Investigation. Council on Library and Information Resources. June 2000. (ISBN 1-887334-78-5). p.15.  
<http://www.clir.org/pubs/abstract/pub93abst.html>
30. A striking example of the potential pitfalls of reliance of converting email to paper can be found in the PROFS case involving the National Archives and Records Administration in the USA.  
[http://www.cpsr.org/cpsr/foia/PROFS\\_CASE/profs.html](http://www.cpsr.org/cpsr/foia/PROFS_CASE/profs.html)

**Update 17-02-2005**

This document is no longer accessible. Please see

<http://www.citizen.org/litigation/briefs/FOIAGovtSec/articles.cfm?ID=873> for information on the PROFS case

31. Dublin Core Metadata Initiative.  
<http://purl.org/DC>
32. National Library of Australia. Draft Preservation Metadata Set. October 1999. p. 2.  
<http://www.nla.gov.au/preserve/pmeta.html>
33. Ross, S. and Gow, A. (1999). Digital Archaeology: Rescuing Neglected and Damaged Data Resources. p. 6.  
<http://www.hatii.arts.gla.ac.uk/research/BrLibrary/rosgowrt.pdf>
34. Lawrence, G.W. et al. (2000). Risk Management of Digital Information: a File Format Investigation. Council on Library and Information Resources. June 2000.  
<http://www.clir.org/pubs/abstract/pub93abst.html>
35. 'Risk Management of Digital Information: a File Format Investigation'. RLG DigiNews 4 (3). 15 June 2000. p.14.  
<http://www.rlg.ac.uk/preserv/diginews/diginews4-3.html>

**Update 07 July 2006**

<http://www.rlg.org/preserv/diginews/diginews4-3.html>

36. Okerson, Ann (1992) 'Synopsis' in Cummings, A. M. et al University Libraries and Scholarly Communication: A Study Prepared for the Andrew W Mellon Foundation,

p.xxiii.

<http://etext.lib.virginia.edu/reports/mellon/>

37. See British Standards Institution (1999). Code of Practice for Legal Admissibility and Evidential Weight of Information Stored Electronically. DISC PD0008: 1999.

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