

Digital Preservation **Handbook**

Organisational Activities



Illustration by Jørgen Stamp digitalbevaring.dk CC BY 2.5 Denmark

Who is it for?

Creators and publishers of digital resources, third-party service providers, operational managers (DigCurV Manager Lens) and staff (DigCurV Practitioner Lens) 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

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 technology on their operational activities.

Gold sponsor



Silver sponsors



Bronze sponsors



Reusing this information

You may re-use this material in English (not including logos) with required acknowledgements free of charge in any format or medium. See [How to use the Handbook](#) for full details of licences and acknowledgements for re-use.

For permission for translation into other languages email: handbook@dpconline.org

Please use this form of citation for the Handbook: Digital Preservation Handbook, 2nd Edition, <http://handbook.dpconline.org/>, Digital Preservation Coalition © 2015.

Contents

Creating digital materials	4
Resources	8
Case studies	11
References	12
Acquisition and Appraisal	13
Resources	21
Case Studies	24
Retention and Review	25
Resources	27
References	28
Storage	28
Resources	34
Case studies	35
References	36
Legacy Media	37
Resources	38
Case studies	40
References	40
Preservation Planning	41
Resources	43
Case studies	44
References	44
Preservation Action.....	45
Resources	47
Case studies	49
References	50
Access.....	50
Resources	53
Case studies	55
Metadata and documentation.....	56
Resources	59
Case studies	61
References	62

Creating digital materials



Illustration by Jørgen Stamp digitalbevaring.dk CC BY 2.5 Denmark

Introduction

"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.

Clearly, most individual creators cannot be expected to take on a long-term commitment to preserving the digital content they create beyond that of their business needs. Every digital resource has a life cycle and different stakeholders and interests within this. However, it is both highly desirable and achievable that a dialogue is established between long-term repositories and creators when issues of long-term preservation are involved. It is often in the creator's interest as well that content created is well-formed, complete, correct, and usable for current and future purposes. Given the crucial role of the creator in undertaking short to medium-term preservation often for a period of decades and at least facilitating medium to long-term preservation, encouraging good practices (and also outreach by repositories), are crucial.

This section will focus solely on encouraging good practices in the creation of digital materials which will assist in their longevity of active use, future management and preservation. You should refer to other relevant sections of the Handbook for related activities and guidance.

Our focus remains the generic implications for digital preservation in the creation process of digitisation (digital surrogates) or that of born-digital materials.

Creating digital surrogates or domain specific types of born-digital files such as electronic records, research data, or personal digital information all have excellent sources of further advice and guidance. Key references are provided in [Resources and case studies](#).

Creating born-digital materials

Digital preservation refers to the series of managed activities necessary to ensure continued access to digital materials for as long as necessary. This includes the activities when creating born-digital materials necessary to meet the ongoing needs of the original creator.

Often many of these actions needed for continuing access for the long-term overlap with those best practices suited to immediate business needs. Indeed many organisations and individuals create digital materials now that they will need to use and manage for many decades. They would probably not consider themselves as doing digital preservation and other terms such as "digital continuity" are frequently used to communicate how these actions affect them when they are not memory organisations such as museums, libraries or archives with a mission to preserve.

It is important for creators to realise if they do not actively work to ensure continuity, their digital materials can easily become unusable. It is about making sure that their information is complete, available and therefore usable for their business needs.

Your information is usable if you can:

- Find it when you need it;
- Open it as you need it;
- Work with it in the way you need to;
- Understand what it is and what it is about;
- Trust that it is what it says it is.

This enables you to operate accountably, legally, effectively and efficiently. It helps you to protect your reputation, make informed decisions, reduce costs, and deliver better services. For further information on first steps in digital preservation see [Getting started](#).

The following table provides guidance on key issues and actions to consider when creating digital materials to ensure their longevity of active use and potential for long-term preservation.

Preserving born-digital materials	
1	<p>Software and formats</p> <p>Choose software that is well supported and creates files that can be read by a variety of different programs.</p> <p>See the File formats and standards section of the Handbook for relevant guidance.</p>
2	<p>File names</p> <p>Use a short descriptive file name of content and date that provide context and can be easily understood by humans and computers, now and in the future.</p> <p>Do not use spaces or special characters (other than - or _), this will avoid potential mis-interpretation by computer hardware or software.</p> <p>Put date information in the ISO 8601:2004 standard format: YYYY-MM-DD. This provides a consistent method for version tracking. Note separate file date metadata generated by systems can often change automatically with later actions.</p> <p>Use a consistent method for showing the file versions. This can be date as above, supplemented as needed , e.g.by a version number v1, v2, v_final, etc.</p>

3	Storage and backup See Principles for using IT storage systems for digital preservation in the Storage section of the Handbook for relevant guidance.
4	Know your obligations and relevant best practice See the Legal compliance section of the Handbook for relevant guidance. Many obligations and best practices will be project or sector specific – see the Creating Research Data inset below for an example.
5	Plan for transitions Some transitions can be foreseen and planned for others may be unforeseen but can be mitigated by good planning and procedures. See Resources and case studies below and the Preservation planning and Risk and change management sections of the Handbook for relevant guidance.

Creating digital surrogates

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 is on the preservation of born-digital materials, or the products of digitisation (the digital surrogates themselves), not the preservation of the analogue originals.

One important exception is that for audiovisual materials. Audio and video materials need digitisation for the very survival of their content, owing to the obsolescence of playback equipment and decay and damage of physical items, whether analogue or digital (see [Moving pictures and sound](#)).

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. It is unlikely that all current digitisation initiatives are being undertaken with due regard to the long-term viability of the digital surrogates they are creating, so it is useful to encourage good practice in creating digital materials and to point to existing sources of guidance.

A good portion of what is now being digitized began life as born digital content. It was converted into an analogue format such as print on paper before the need for digital access and to re-digitize was recognised. That cycle needs to shift quickly to simply managing more born digital content.

Preserving digital surrogates: digital preservation considerations	
1	Assessment of need for digitisation Has the material already been digitised? If so, is it to an appropriate standard and readily accessible to your audience?
2	Finding funds for the project

	<p>What archiving policies exist, both from the funding agency (if externally funded) and the institution with prime responsibility for the project?</p>
3	<p>Planning the project and assigning resources</p> <p>Need to set aside recurrent funds for maintenance of the digital copies as well as one-off funds for conversion.</p> <hr/> <p>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)</p> <p>Identify a strategy for carrying forward the assets of the project in a sustainable manner after the project has achieved its deliverables. This strategy might involve ingesting the assets of the project into the collection catalogue of the parent organisation, or designating a partner institution for receipt of these assets.</p>
4	<p>Selection of materials</p> <p>Copyright. It will be necessary to ensure permission is given both to digitise the original and to make copies of the digital copy for the purposes of preservation and delivery. For further information, see Legal compliance.</p> <hr/> <p>Condition and completeness of original. Is it capable of being re-scanned at a later date if the digital copy is lost?</p>
5	<p>Decide how the information content needs to be organised</p> <p>(for example, searchable text databases and/or document page images)</p> <p>Selection of appropriate file formats and storage for both master/archive copies and derivatives, see File formats and standards, Metadata and documentation and Storage.</p>
6	<p>Decide digitisation method appropriate to analogue original and goals of the project.</p> <p>Preparing originals for digitisation Details of the digitisation method need to be documented and attached to the metadata record to enable future management.</p>
7	<p>Preparing originals for digitisation</p> <p>The National Archives provides standards and guidance on document preparation for digitisation of records (The National Archives, 2015).</p> <p>Will the originals be retained? Do 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.</p> <p>Deciding whether or not to retain the originals post-digitisation 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</p>

	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. <hr/> 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.
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.

Resources



Digitisation at The National Archives

<http://nationalarchives.gov.uk/documents/information-management/digitisation-at-the-national-archives.pdf>

This document sets out TNA's standards and requirements for the digitisation of analogue records in its collection. It is also recommended to UK government departments who wish to digitise any of their paper records. It covers: the whole digitisation process from initial scanning through to delivery of the images for preservation, including The National Archives' scanned image specification; the scanning of records where the resultant images will become the legal public record for permanent preservation; and the scanning of records where the resultant images will become digital surrogates with the original paper records being retained and remaining the legal public record (July 2015, 56 pages).

Koninklijke Bibliotheek/National Library of the Netherlands: Metamorfoze preservation imaging guidelines

http://www.metamorfoze.nl/sites/metamorfoze.nl/files/publicatie_documenten/Metamorfoze_Preservation_Imaging_Guidelines_1.0.pdf

Metamorfoze is the national program of the Netherlands for preserving paper heritage. The guidelines are intended for the digitisation of two-dimensional materials such as manuscripts, archives, books, newspapers and magazines. They may also be applied to photographs, paintings and technical drawings. The Guidelines relate exclusively to the image quality and metadata of the Preservation Master file, from which all outputs intended for print and/or the web can be derived. (2012, 44 pages).

Preparing Collections for Digitisation

This 2010 book by Anna E. Bulow and Jess Ahmon offers practical guidance covering the end-to-end process of digitising collections, and can be used as a 'how-to' reference manual for collection managers who are embarking on a digitisation project or who are managing an existing project. It also covers some of the wider issues such as the use of surrogates for preservation, and the long term sustainability of digital access. (208 pages).

InterPARES 2 Creator Guidelines Making and Maintaining Digital Materials

[http://www.interpares.org/ip2/display_file.cfm?doc=ip2\(pub\)creator_guidelines_booklet.pdf](http://www.interpares.org/ip2/display_file.cfm?doc=ip2(pub)creator_guidelines_booklet.pdf)

This booklet provides advice for individuals who create digital materials in the course of their professional and personal activities to help them ensure their preservation (10 pages).

InterPARES 2 Preserver Guidelines Preserving Digital Records: Guidelines for Organisations

[http://www.interpares.org/public_documents/ip2\(pub\)preserver_guidelines_booklet.pdf](http://www.interpares.org/public_documents/ip2(pub)preserver_guidelines_booklet.pdf)

This booklet provides advice to any organization responsible for the long-term preservation of digital records (10 pages).



Jisc Digital Media resources

Copyright and still images: Frequently Asked Questions

<http://www.jiscdigitalmedia.ac.uk/guide/copyright-and-still-images-frequently-asked-questions>

infokit: Digitisation funding and sustainability

<http://www.jiscdigitalmedia.ac.uk/infokit/digitisation-funding-and-sustaina/digitisation-funding-and-sustaina-home>

infokit: High level digitisation for audiovisual resources

<http://www.jiscdigitalmedia.ac.uk/infokit/audiovisual-digitisation/audiovisual-digitisation-home>

infokit: Still image digitisation

<http://www.jiscdigitalmedia.ac.uk/infokit/digitisation/still-image-digitisation-home>

infokit: Digital file formats

http://www.jiscdigitalmedia.ac.uk/infokit/file_formats/digital-file-formats

Federal Agencies Digitization Guidelines Initiative

<http://www.digitizationguidelines.gov/>

This is a collaborative effort by US federal agencies to define common guidelines, methods, and practices for digitizing historical content. As part of this, two working groups are studying issues specific to two major areas, Still Image and Audio Visual.

Future Proof – Protecting our digital future

<http://futureproof.records.nsw.gov.au/>

Future Proof is a State Records initiative from the New South Wales State Government in Australia. This website and blog cover products and projects from State Records that are specifically about digital records. Category links provide useful gateways into different resources housed on the site.

The Curation Reference Manual

<http://www.dcc.ac.uk/resources/curation-reference-manual>

This resource maintained by the Digital Curation Centre contains advice, in-depth information and criticism on current digital curation techniques and best practice. The Manual is an ongoing, community-driven project, which involves members of the DCC community suggesting topics, authoring manual instalments and conducting peer reviews. Each instalment is designed to help data custodians, producers and users better understand the challenges they face and the roles that they play in creating, managing and preserving digital information over time.

UK Data Archive: Create and Manage Data - Ethical / Legal / Overview

<http://www.data-archive.ac.uk/create-manage/consent-ethics/legal>

Collecting, using and sharing data in research with people requires that ethical and legal obligations are respected. Laws such as the Data Protection Act, Freedom of Information Act and Statistics and Registration Services Act also govern the use of some kinds of data. This guidance offers help on how research data can be shared without breaching ethical or legal responsibilities.

An Elevator Pitch for File Naming Conventions

<http://acrl.ala.org/techconnect/post/an-elevator-pitch-for-file-naming-conventions>

This Association of College and Research Libraries (ACRL) TechConnect blog post makes the case for adopting a consistent approach when naming digital files or software components, by demonstrating the effects of not doing so. (2013).

Digital Continuity guidance from The National Archives UK

<http://www.nationalarchives.gov.uk/information-management/manage-information/policy-process/digital-continuity/>

Comprehensive guidance on digital continuity from the National Archives. Of particular use are: [Understanding Digital Continuity](#) an introduction to the topic (2011, 20 pages); and [Managing Digital](#)

[Continuity](#) which takes you through a 4 stage process:1) Plan for action 2) Define your digital continuity requirements 3) Assess and manage risks to digital continuity 4) Maintain digital continuity.

NCDCCR Digital Preservation Best Practices and Guidelines - Create Digital Files

<http://digitalpreservation.ncdcr.gov/>

First launched in 2010 by the Digital Information Management Program of the State Library of North Carolina, and the Digital Services Section of the State Archives of North Carolina, this site received a National Digital Stewardship Alliance Innovation Award in 2012. The aim is to provide practical, introductory information about digital preservation, and to direct visitors to approachable "next step" resources.

Digital Preservation Management Tools and Techniques

<http://dpworkshop.org/workshops/management-tools>

DPM workshop content workflows include a high-level diagram with lower-level diagrams for managing physical content, transitioning through digitisation, and managing born-digital and digitised content. The idea is to provide a common workflow for all content in any context then develop a number of use cases to highlight exceptions for specific kinds of content with different kinds of requirements.



Part 1: Why is File Naming Important?

https://www.youtube.com/watch?v=Hi_A4Ywn4VU

This excellent short video is part one of a four-part tutorial on file naming. It talks about why it's important to choose your file names wisely. Designed for a general audience, it is part of the State Library of North Carolina's "Inform U" series (2012, 3mins 19 secs).

Case studies



DPC case note: ULCC assessing long term access from short term digitization projects

http://www.dpconline.org/component/docman/doc_download/534-casenoteassessingpreservationindigitization.pdf

Digitisation projects are mostly funded over a short term, so how can we take steps to make the outputs of digitisation robust in the long term? This Jisc-funded case study reports work undertaken by the University of London Computer Centre in assessing the long term plans of 16 digitisation projects, providing a basic survey tool to help funders and project managers alike to reflect on their long term preservation plans. November 2010 (4 pages).

The British Library 'Save our Sounds' project

<http://www.bl.uk/projects/save-our-sounds>

Launched in 2015 Save our Sounds is the British Library's programme to preserve via digitisation the nation's Sound Archive, a collection of over 6.5 million recordings of speech, music, wildlife and the environment, from the 1880s to the present day. The project aims both to ensure that the existing archive is properly preserved, and that there are adequate systems in place for the acquisition of future sound production in the UK.

Digital Curation Centre case studies

<http://www.dcc.ac.uk/resources>

In 2013 the DCC began a series of case studies to accompany the new DCC guide How to Develop Research Data Management Services. These cover specific components of a Research Data Management service of interest to researchers and data managers.

Society of American Archivists campus case studies

<http://www2.archivists.org/publications/epubs/Campus-Case-Studies>

Campus Case Studies are reports by American university archivists who have created working solutions. They cover a wide range of topics some of which are specifically focussed on digital preservation and creating digital records. The currency of the case studies varies from 2008 to the present.

Why metadata matters

<https://cbaileymsls.wordpress.com/2013/09/29/metadata/>

This blog post provides good examples of why poor file-naming and metadata description at creation of a file can hinder subsequent searching, discovery and re-use.

References

The National Archives, 2015. *Digitisation at The National Archives*. Available:

<http://nationalarchives.gov.uk/documents/information-management/digitisation-at-the-national-archives.pdf>

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. Available:

<http://www.oclc.org/programs/ourwork/past/digpresstudy/final-report.pdf>

Acquisition and Appraisal



Illustration by Jørgen Stamp digitalbevaring.dk CC BY 2.5 Denmark

Introduction

In a digital environment, decisions taken regarding 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 may need to 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 physical dimensions of an archive mean the reasons to select are relatively well understood, while the decision to preserve can be taken quite separately and within a timeframe which may span several decades. In contrast, digital resources can proliferate and appraisal can be a daunting task. Moreover they can become inaccessible relatively quickly, so decisions about selection and preservation may need to be taken simultaneously for digital collections.

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 later as retrospective preservation of digital resources is not recommended.

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 hyperlink to a resource, online catalogues and finding aids.

In some cases an institution may be reluctant to take primary preservation responsibility for material 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. A number of services have emerged in recent years, like the [Keepers Registry](#) for electronic journals or the [DLF/OCLC Registry of Digital Masters](#) which allow institutions to identify preservation intent - their commitment to preserve material that may be of general interest. 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 choices may need to be made about whether the resource justifies the costs or whether it is preferable not to proceed with acquisition.

Development of policy and procedure

Before embarking on the acquisition and ingest of digital collections it may be necessary to establish whether current policies (e.g. collection development) and procedures are still fit for purpose in the digital age (see [Institutional policies and strategies](#)). Depending on the structure and wording of existing documents this review may result in anything from small changes that increase scope to include digital objects through to the creation of new policy documents specifically covering digital collections. Additions or alterations to the policy may include descriptions of the types of objects that will be acquired, in relation to format and/or content, as well as addressing other issues such as intellectual property rights, sensitivity and access considerations. It is essential that any changes are ratified by the relevant management committees within your organisation to ensure support and buy-in.

Appraisal/retention policy

The [Decision Tree](#) accompanying this section may be used as a tool to construct or test the selection, or appraisal/retention policy for your organisation.

Appraisal of born digital objects should include a measured assessment of their value to the parent organisation set against the challenges of long-term preservation and providing access. These challenges may include an organisation's ability to read or open a version of the master file, the ability to secure sufficient rights to manage and provide access to current and future versions of the file, or simply staffing and funding resources. Organisations should therefore initially focus on a balance between acquisition of high value digital objects and these longer term curatorial obligations. It should be remembered that organisations can provide access to resources that they have accessioned without placing them in specific preservation or retention workflows. A detailed policy document which clearly identifies the most important digital resources (from either a format or content perspective) can give guidance on appraisal of born digital objects destined for such pathways. For lesser value digital acquisitions, which often come bundled with higher value acquisitions, it may be enough to outline the level of access and preservation an organisation will provide to them. This outline should include an indication of a retention schedule suitable to this type of content. This may mean also including a disposal schedule or de-accessioning policy if appropriate (see [Retention and review](#)).

Agreements and Guidance for depositors - file formats, required documentation

Once policy has been established there will be a number of additional supporting documents that will be required to facilitate the acquisition and appraisal process. Alongside the standard procedural documents an organisation may wish to create a suite of standard depositor agreements and licences to aid in the negotiation process. These will be particularly useful in ensuring that the minimum permissions and intellectual property rights required for preservation are granted. Without sufficient licence agreements an organisation may find itself in possession of digital collections that it does not hold the rights to actively preserve or provide access to (see [Legal compliance](#)). These may also be complemented by guidance notes for depositors that set out requirements for material to be transferred and accompanying documentation.

Standards for acquisition and transfer

Experience shows that the transfer from a producer to an archive can be tortuous and therefore any tools which can streamline the process are likely to benefit both sides. Two initiatives have attempted to standardise the interface between Producers and Archives into a consistent, well-understood process, cultivating a mutual understanding between producers and archives in regard to their respective roles: Producer-Archive Interface – Methodology Abstract Standard (PAIMAS, ISO 20652:2006); and the Producer–Archive Interface Standard (PAIS, ISO 20104:2015).

PAIMAS provides a standardized description of the interactions between producers and an archive. It segments the transfer process into a number of phases, providing a detailed description of the anticipated outcome of each phase and the actions required to bring about this outcome. The four principle phases - Preliminary phase, Formal Definition Phase, Transfer Phase, Validation Phase - serve as a basis for identifying areas within the Producer-Archive interface that would benefit from more focused standards, recommendations, and best practices, and also provides a foundation for the development of automated processes and software tools to support the information transfer process. PAIMAS implicitly expands the detailed requirements for Ingest and Administration within the OAIS reference model.

PAIS provides a standard method for formally defining the digital information objects to be transferred by an information Producer to an Archive and for effectively packaging these objects in the form of Submission Information Packages. It is intended to support more precise definitions of the digital objects, helping archives process and validate objects received during submission.

Acquisition workflow

Negotiation

Negotiating the terms of the deposit should take place before any records have been transferred. Many aspects of the deposit agreement may be covered in the acquisition policy of the organisation but details about each deposit, especially for local and specialist archives may be required at a collection level. The depositor should state if there are any limitations in what or when the records can be published, for example can some material be opened immediately while others can only be opened on the death of the depositor or after a set period of time?

A key consideration is the right of the organisation to alter the record for preservation purposes, for example migration to a format that can be preserved in the long term or accessed.

If the transfer includes content that is essential to the understanding of the records but does not constitute a record itself there should be an agreement that the organisation can delete those files when their content has been captured for use elsewhere (for example as metadata for the records).

Transfer

Most institutions will need to develop procedures and documents to support the smooth transfer of digital resources from suppliers into their collections.

When transferring born digital objects into an organisation's IT environment consideration should be given as to how this is to take place so as to ensure the security and completeness of the transfer. For smaller organisations it may be sufficient to have the relevant digital files delivered on a drive or similar hardware and check their contents against the descriptive file manifest. Alternatively organisations may wish to transfer digital files using an in-house FTP, or a paid for third party solution (such as a cloud based file sharing service), to ensure chain of custody.

The table below outlines options for transfer and accessioning of digital materials. Decisions on file formats and if relevant storage media (see [Storage](#), [Legacy media](#), and [File formats and standards](#)) will support and be interdependent with this process.

Options for Transfer and Accessioning of File Formats and Storage Media		
Options	Issue	Requirements
All options		<ul style="list-style-type: none"> • Policy on File formats and standards. • Preservation planning and technology watch on developments in Storage and for Legacy media formats.
Limit range of file formats received Limit range of media received (most cost-effective long-term option)	<ul style="list-style-type: none"> • Simplifies management and reduces overall costs. • Depositor may lack resource or expertise to comply. • Wide variety of file formats used and proprietary extensions to open standards. • Physical storage media used for transfer may only be temporary carriers and content will be transferred to long-term storage used. 	<ul style="list-style-type: none"> • Guidelines on preferred file formats. • Degree of influence over the deposit. • Advocacy and Collaboration strategies to achieve desired outcomes. • Guidelines on preferred transfer media and transfer procedures.
Accept file formats as received but convert to standard file format Accept storage media as received but transfer contents to standard storage used	<ul style="list-style-type: none"> • Simplifies management and reduces longer term costs. • May not be technically feasible to convert to standard file format. • It will be necessary to check that accidental loss of data has not occurred. 	<ul style="list-style-type: none"> • Legal compliance, Copyright permissions or statutory preservation rights. • Resources and technical expertise at host institution. • Election of preferred formats.

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

Validation

Once transfer has taken place, the files should be located in a secure, quarantined, backed up environment and a check in process should be promptly initiated. Having transferred and housed a copy of a born digital collection an organisation is now liable for certain legal responsibilities such as Freedom of Information Requests if in the public sector. Following this, a letter of acknowledgement of receipt should be sent to the donor. It is important at this early stage that no instruction to destroy the original files is given.

Records should be virus checked at the earliest opportunity to ensure that the material has not been infected with malware or viruses. If any are found the depositing organisation should be alerted and the media either returned to them (if they do not have a copy) or formatted and returned or destroyed according to the depositor's preference. Once the records are confirmed as virus free a check should be carried out to ensure that all the records are present and undamaged. The most reliable method for this is to verify the files against the manifest. Create checksums for the files and compare them against those listed on the manifest pre-transfer. If the checksums match you can be sure that the records have not been corrupted or accidentally altered between the points of transfer from the depositor and arrival at the organisation. If no verifiable manifest was provided with the deposit, it may be impossible to comprehensively verify the integrity of the files and manual viewing of a sample of files may be necessary to provide some indication of completeness and quality. In this case, a verifiable manifest should be generated to enable subsequent fixity checking.

At this stage the records will hopefully have been confirmed as complete (according to the manifest), retaining their integrity (exactly what the depositor supplied) and are virus and malware free. They can now be ingested into the digital preservation system.

Metadata describing the deposited material will assist in ensuring the fixity (see Fixity and checksums) of the material during the transfer process as well as supporting subsequent preservation and access. This might include:

A verifiable manifest consisting of a list of the file and folder names and checksums/fixity values for each file

The size of the files (with a total volume)

A list of the file formats

A statement detailing any IPR associated with the records

Where possible the onus of providing information on the IPR of the records should reside with the depositor.

Ingest Process

The period between transfer to the organisation and ingest into the organisation repository or digital preservation environment may be substantial. This accession phase can be especially prolonged for large born digital collections, sometimes amounting to years, but it is during this phase that a qualitative appraisal of the objects can be made. Items are examined, their technical metadata harvested, their descriptive metadata enhanced and the general accession processes of the organisation applicable to any object take over.

It is during this sometimes prolonged appraisal period that items can be reconsidered for ingest, or rejected if on examination it is felt they do not meet the acquisition or collection profile of the organisation, the file format specification laid out in the guidance documents, or for any other reason. A moratorium may be imposed on items of particular sensitivity such as personal information, commercially sensitive information, or items that break libel laws for instance. In such cases it is important to clearly specify the closure period of the file.

Ingest Procedures to prepare data and documentation for storage and preservation

Unique numbering

Each digital 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. See [Persistent identifiers](#) for advice if you use a persistent identifier scheme.

Handling and Storage transfer guidelines

Handling and transfer guidelines for accessioning staff should be developed reflecting IT and preservation staff advice on best practice for different storage media and file transfer to long-term storage systems (see [Legacy media](#), [Digital forensics](#), and [Storage](#)).

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 [File formats and standards](#), and [Storage](#)).

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](#)).

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 [Information security](#)).

Edition and version control

Procedures for updating and edition control of any dynamic digital materials 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 have been generated.)

Cataloguing and documentation standards

Metadata and documentation received or created during transfer, validation and ingest 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), content, structure, and about the terms and conditions attached to its subsequent management and use including IPR rights and the period over which they pertain (see [Metadata and documentation](#)). 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.

Skills, resources and capacity

Organisations should consider whether they have sufficient technical and staffing resources to acquire digital collections. This information may however not be apparent at the outset of an acquisition, as the various challenges of curation of specific digital collections may only reveal themselves over time. Organisations should therefore plan for knowledge, skill and staffing gaps and where possible address these through training, recruitment or engagement of specialist professional digital curation services. Where funding resources cannot meet these, often a dedicated in-house knowledge building drive may suffice for the interim. (see [Staff training and development](#), and [Procurement and third party services](#)).

Costs of acquisition and ingest

Trying to establish indicative costs for digital preservation activities is always problematic. These should not just include storage (the most obvious) but should also look at the cost of the staff time required to manage the accession and ingest of each born digital object, a process which can mirror time-wise the accession pathway of physical artefacts. Other anticipated costs might include curation processes like normalisation, analysis, enrichment of metadata, increased robustness of storage, disaster recovery etc.

Although an organisation should find best value solutions for these lifecycle costs, it should be recognised that the investment needed to provide a robust preservation pathway that can safeguard our digital heritage may be significant, and that certain processes must be adhered to irrespective of the nature of an accessioned object. Organisations should therefore bear this in mind when acquiring born digital collections. (see [Business cases, benefits, costs, and impact](#))

Summary of recommendations

Acquisition and appraisal - recommendations checklist

Agreements and Guidance for depositors

- Create a suite of standard depositor agreements and licences
- Create appropriate guidance for depositors

Transfer procedures

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

Validation procedures

- Check media, content, and structure

Procedures to prepare data and documentation for storage and preservation

- Unique numbering of each item accessioned
- Handling and storage transfer guidelines for different media
- Re-formatting of file formats if required 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 including IPR information
- 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 standards
- Targets for accessioning tasks and timescales for their completion

Review of procedures

- Guidelines and schedules should ideally be reviewed annually, or as often as is practical to keep pace with an organisation's developing requirements and collections development policies

Staff training

- Plan for knowledge, skill and staffing gaps and where possible address these through training, recruitment or engagement of specialist third-party services

Costs

- Evaluate and plan for lifecycle costs of acquisitions

Resources



ISO 20104:2015 Space data and information transfer systems -- Producer-Archive Interface Specification (PAIS)

CCSDS 651.1-B-1, Producer-Archive Interface Specification (PAIS) (2014) RECOMMENDED STANDARD CCSDS 651.1-B-1 BLUE BOOK February 2014

<http://public.ccsds.org/publications/archive/651x1b1.pdf>

The Blue Book is a free to access pre-print of ISO 20104:2015. The PAIS standard aims to provide a standard method for formally defining the digital information objects to be transferred by an information Producer to an Archive and for effectively packaging these objects in the form of Submission Information Packages (SIPs). This supports effective transfer and validation of SIP data (104 pages).

What is appraisal?

<http://www.nationalarchives.gov.uk/documents/information-management/what-is-appraisal.pdf>

This guidance from The National Archives applies to UK public records in any format, including paper, digital, audio, film or model format as defined by the Public Records Act 1958, and all organisations responsible for such records (2013, 7 pages).

Preserving eBooks, DPC Technology Watch Report 14-01 July 2014

<http://dx.doi.org/10.7207/twr14-01>

This report discusses current developments and issues with which public, national, and higher-education libraries, publishers, aggregators, and preservation institutions must contend to ensure long-term access to eBook content and which affect acquisition as well as preservation (31 pages).

Preservation, Trust and Continuing Access for e-Journals, DPC Technology Watch Report 13-04 September 2013

<http://dx.doi.org/10.7207/twr13-04>

This report discusses current developments and issues which libraries, publishers, intermediaries and service providers are facing in the area of digital preservation, trust and continuing access for e-journals. It is not solely focused on technology, and covers relevant legal, economic and service issues in acquiring access to networked digital resources and the unique preservation challenges this presents (43 pages).

The UNESCO/PERSIST Guidelines for the selection of digital heritage for longterm preservation

<https://www.unesco.nl/sites/default/files/dossier/persistcontentguidelinesfinal1march2016.pdf?download=1>

The UNESCO/PERSIST (Platform to Enhance the Sustainability of the Information Society Transglobally) Project released these Guidelines on the selection of digital heritage for long-term preservation in March 2016. The aim of the Guidelines is to provide an overarching starting point for libraries, archives, museums and other heritage institutions when drafting their own policies on the selection of digital heritage for long-term sustainable digital preservation. (19 pages).



Community Owned digital Preservation Tool Registry COPTR

http://coptr.digipres.org/Main_Page

COPTR describes tools useful for long term digital preservation and acts primarily as a finding and evaluation tool to help practitioners find the tools they need to preserve digital data. COPTR captures basic, factual details about a tool, what it does, how to find more information (relevant URLs) and references to user experiences with the tool. The scope is a broad interpretation of the term "digital preservation". In other words, if a tool is useful in performing a digital preservation function such as those described in the OAIS model or the DCC lifecycle model, then it's within scope of this registry. You can use the [POWRR Tool Grid](#) to see which technical tools in COPTR can help to support acquisition, ingest, or multiple functions.

DLF/OCLC Registry of Digital Masters

<http://www.diglib.org/community/groups/rdm/>

The DLF/OCLC Registry of Digital Masters provides a central place for library staff to search for, and find, digitally preserved materials. Typical items include digitised monographs and serials. A registered object ensures that the digital object (or soon to be digitised) followed established standards and best practices for digitisation and that the institution that digitised it has made a commitment to digital preservation of this object.

Keepers Registry

<http://thekeepers.org>

The Keepers Registry acts as a global monitor on the archiving arrangements for electronic journals. It has three main purposes to: enable librarians and policy makers to find out who is looking after which e-journal, how and with what terms of access.; highlight the e-journals which are still "at risk of loss"; and showcase the organisations (the keepers) which act as digital shelves for access over the long term. It has a Title List Comparison feature to help you discover the archival status of a list of serial titles important to you: reporting those which are being archived and those which are "at risk".

MediaRIVERS (Media Research and Instructional Value Evaluation and Ranking System)

<https://github.com/IUMDPI/MediaSCORE>

Software created by Indiana University in collaboration with AVPreserve guides a structured assessment of research and instructional value for media holdings. The free, open source version requires installation and configuration on a server, and a hosted application is available on a monthly subscription basis.

Practical E-Records:software and tools for archivists

<http://e-records.chrisprom.com/>

Pages created by Chris Prom for Transfer Guidelines, E-Records Deposit Policy, and Submission Agreement Form provide sample templates that you can modify and/or provide to record producers whose records your repository wishes to accession. Permission to modify and republish these transfer guidelines is provided under a Creative Commons Attribution 3.0 United States License.



Archaeology Data Service Guidelines for Depositors

<http://archaeologydataservice.ac.uk/advice/guidelinesForDepositors>

The ADS Guidelines for Depositors provide guidance on how to correctly prepare data and compile metadata for deposition with ADS and describe the ways in which data can be deposited. There is also a series of shorter summary worksheets and checklists covering: data management; selection and retention; preferred file formats and metadata. Other resources for the use of potential depositors include a series of Guides to Good Practice, which complement the ADS Guidelines and provide more detailed information on specific data types.

Selecting and transferring records

<http://www.nationalarchives.gov.uk/information-management/manage-information/selection-and-transfer/>

These pages provide guidance on the selection and transfer of records. UK bodies transferring records to The National Archives or to places of deposit under the Public Records Act 1958 should follow this process for records in all formats and media, including paper and digital records. It consists of guidance on six steps:

Step 1: Appraising your records

Step 2: Selecting your records

Step 3: Sensitivity reviews of selected records

Step 4: Cataloguing and preparation of records

Step 5: Planning and arranging delivery of records

Step 6: Accessioning your records

Case Studies



The Work of Appraisal in the Age of Digital Reproduction

<http://archival-integration.blogspot.co.uk/2015/06/the-work-of-appraisal-in-age-of-digital.html#pii>

The Bentley Historical Library's ArchivesSpace-Archivematica-DSpace Workflow Integration project discussion highlights current digital archives appraisal techniques employed by the Bentley, many of which they are hoping to integrate into Archivematica (June 2015).

Acquisition & management of digital collections at the Library of Congress

<http://www.slideshare.net/NASIG/acquisition-management-of-digital-collections-at-the-library-of-congress-34244613>

The Library of Congress, as the national library and the home of the US Copyright Office, is heavily involved in digital acquisition and management. This concise and informative powerpoint by Ted Westervelt shares the experiences that the Library of Congress has had and lessons it has learned (2014, 30 slides).

Trust Me, I'm an Archivist: Experiences with Digital Donors

<http://www.ariadne.ac.uk/issue65/hilton-et-al>

This 2010 article by staff at the Wellcome Trust Library discusses four common scenarios that seem to act as new blocks to the transfer of digital material: Lack of Long-term Planning; IT vs Records Management; Duplication and Abundance; and The Fear of Digital. It concludes that we need to change the way we present information, how we work with digital material and how we can support and assist our donors. The degree of engagement that is standard practice with paper records will not suffice for born-digital material: our interaction with depositors will ideally be even closer and even more frequent, as we help them deal not merely with new technical challenges but with the plethora of soft-skills issues, of preconceptions and of attachments that surround them.

Retention and Review



Illustration by Jørgen Stamp digitalbevaring.dk CC BY 2.5 Denmark

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 [Acquisition and appraisal](#) 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. Organisations may also need to retain certain in house records and digital materials for regulatory, legal, operational and financial requirements. These should similarly be actively managed to retain their viability, authenticity, and accessibility.

Digital items acquired over time and before institutional policies and procedures were in place will normally require 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 [Getting started](#)).

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 digital materials.

In a digital library environment where collection levels have been employed, digital materials 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 the 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 with clear responsibilities assigned to named individuals or those sections of an organisation in charge of governance and collections development. The timescales, circumstances, and authorisation procedures for the review should be clearly stated. Retention and review schedule documents themselves should be periodically reviewed to keep pace with emerging organisational requirements. 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.

Financial constraints should not be a main driver for de-accessioning digital objects as the process itself is not without cost and may raise thorny ethical issues. Decisions to de-accession should primarily be driven by the collections development policy.

Accessioned digital materials 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.

Succession planning by an institution may also be relevant to retention and review. The standard for Audit and Certification of Trustworthy Digital Repositories ([CCSDS, 2011](#)) recommended practice 3.1.2.1 requires that the repository shall have an appropriate succession plan, contingency plans, and/or escrow arrangements in place in case the repository ceases to operate or the governing or funding institution substantially changes its scope.

Resources



Deaccessioning and disposal: Guidance for archive services

<http://www.nationalarchives.gov.uk/documents/Deaccessioning-and-disposal-guide.pdf>

This guide was written to support Archive Service Accreditation, the UK-wide standard for archive services. It is generic and applies to content in all formats, analogue or digital and has no special provisions for digital materials. The standard presents de-accessioning as part of collections development, and requires archive services to have policies, plans and procedures in place for collections development activities including de-accessioning. It includes a disposal destination decision tree (2015, 34 pages).



Digital Preservation Management tools: Digital Content Review process

<http://dpworkshop.org/workshops/management-tools/process-results>

To complete a digital content review, the digital preservation team gathers information and iteratively accumulates as part of a structured process. The results of ongoing digital content reviews produce a digital content review dataset that enables near-term and long-term planning by organizations.

The National Archives, Disposing of records

<http://www.nationalarchives.gov.uk/information-management/manage-information/policy-process/disposal/>

TNA have produced step-by-step guidance to help you through the disposal process including a disposal checklist.



The British Library's de-accessioning policy

<http://www.bl.uk/aboutus/stratpolprog/coldevpol/deaccessioning/>

This policy sets out the circumstances under which the British Library may dispose of certain types of material. It is a generic policy that applies to content in all formats analogue or digital and has no special provisions for digital materials.

References

Consultative Committee for Space Data Systems (CCSDS), 2011. *Audit and Certification of Trustworthy Digital Repositories Recommended Practices CCSDS 652.0-M-1 Magenta Book September 2011* [this was subsequently published as ISO 16363: 2012]. Available: <http://public.ccsds.org/publications/archive/652x0m1.pdf>

Storage



Illustration by Jørgen Stamp digitalbevaring.dk CC BY 2.5 Denmark

Introduction

This section covers the emerging practice of using IT storage systems for digital preservation. It deals with generic issues and more specific issues associated with cloud storage are addressed separately (see [Cloud services](#) section). The traditional practice of preserving of digital media, for example legacy items within existing collections is also covered elsewhere (see [Legacy media](#) section). Many organisations will have mixed strategies or will be in the process of transitioning from one to the other.

The use of storage technology for digital preservation has changed dramatically over the last twenty years. During this time, there has been a change in practice. Previously, the norm was for storing digital materials using discrete media items, e.g. individual CDs, tapes, etc., which are then migrated periodically to address degradation and obsolescence. Today, it has become more common practice

to use resilient IT storage systems for the increasingly large volumes of digital material that needs to be preserved, and perhaps more importantly, that needs to be easily and quickly retrievable in a culture of online access. In this way, digital material has become decoupled from the underlying mechanism of its storage. With this come consequent benefits of allowing different preservation activities to be handled independently.

Resilient storage systems

A resilient IT storage system consists of storage media contained within a server that provides built in resilience to various failure modes by using inbuilt redundancy and recovery. For example, a storage system might be hard disk drives in a Redundant Array of Independent Disks (RAID), data tapes in a tape library, or a combination of storage types in a Hierarchical Storage Management system (HSM). It can include onsite storage and/or remote cloud storage and automated replication of digital materials across multiple sites and systems.

These systems will still become obsolete over time and digital materials should be migrated regularly between storage systems as they become obsolete. Migration between storage systems is separate to migration between file formats and can be handled largely as an IT issue, with the proviso that proper oversight is employed to ensure preservation requirements are met. The upside is that the use of IT systems for data storage can provide much faster access, a more scalable solution, easier management, and ultimately lower costs especially at scale.

It is critical to understand the difference between standard IT storage solutions and the additional needs of long-term preservation. It is essential to be able to explain these differences to your IT department or storage service provider and to be able to specify these requirements when procuring a system or service. Standard storage systems are designed for digital objects that are in active use. While backup procedures are usually included, they generally do not meet the more stringent requirements to ensure long-term preservation of digital materials. Backup and digital preservation are not the same thing and many IT departments or experts may not appreciate this. Preservation storage systems require a higher level of geographic redundancy, stronger disaster recovery, longer-term planning, and most importantly active monitoring of data integrity in order to detect unwanted changes such as file corruption or loss.

There are many ways of meeting requirements for preservation storage and these will vary in scale and complexity depending on organisational context. It will be necessary to assess in-house resources and consider out-sourcing and cloud storage options. The approach taken will often depend on the collection size and complexity of collection and the resources that are available within the organisation. It is possible to meet preservation storage requirements with a basic set-up, but as a collection increases in size it will be necessary to address issues such as scalability and automation.

Principles for using IT storage systems for digital preservation	
The following represent principles that should be employed when designing or selecting storage systems for preservation.	
1	Redundancy and diversity <ul style="list-style-type: none">• Make multiple independent copies of digital material and store these in different geographic locations.

	<ul style="list-style-type: none"> • Use a combination of online storage systems and offline media. • Use different types of storage technology to spread risk and achieve a balance of data safety, easy access and manageable cost.
2	<p>Fixity, monitoring, repair</p> <ul style="list-style-type: none"> • Use fixity measures such as checksums to record and regularly monitor the integrity of each copy of the digital material. • If corruption or loss is detected then use one of the other copies to create a replacement. • Store fixity information alongside the digital materials and also in separate databases or systems.
3	<p>Technology and vendor watch, risk assessment, and proactive migrations</p> <ul style="list-style-type: none"> • Understand that storage technologies, products and services all have a short lifetime. • Use technology watch to assess when migrations might be needed. • Keep an eye on the viability of storage vendors or classes of storage solution. • Be proactive in migrating storage before digital material becomes at risk
4	<p>Consolidation, simplicity, documentation, provenance and audit trails</p> <ul style="list-style-type: none"> • Minimise the proliferation of legacy media types and storage systems used for preservation. • Consolidate digital materials onto the minimum number of preservation storage systems (subject to the redundancy requirements above). • Document how digital materials have been acquired and transferred into the storage systems as well as how the storage systems are set-up and operated. • Use this to provide audit information on data authenticity.

Storage reliability

When looking at storage solutions, either onsite or cloud, the question arises of how reliable they are and if something goes wrong then what does this mean in terms of data loss. Manufacturers will typically assert statistics such as reliability, durability, failure rates and error rates.

For example, this might take the form of a cloud service being designed for 99.999% durability, a Bit Error Rate (BER) of 1 in 10^{16} when reading data from a data tape, or a Mean Time Between Failure (MTBF) of 1M hours for a hard drive. These numbers are then used to calculate further measures of reliability. For example, a Mean Time To Data Loss (MTTDL) of 1,000 years might be asserted when hard drives are used in a RAID6 array.

These numbers can be hard to understand, and they need to be interpreted with great care when attempting to estimate 'how safe' a given storage solution will be.

There has already been substantial work on how to describe, measure and predict storage reliability, including from a digital preservation perspective. This is a complex topic that is not possible to cover

in this handbook. Some example references for further reading are [Greenan et al \(2010\)](#), [Rosenthal \(2010\)](#) and [Elerath \(2009\)](#). What comes from this work are several important considerations:

- IT Storage technology is in general remarkably reliable for what it does. Failures are relatively rare events, but they do and will happen. The temptation is to assume that just because at an individual level a particular type of failure hasn't been experienced then that storage technology is in general more reliable than it really is. This is dangerous position. For example, many people will have hard drives that have worked perfectly well for years and years, but the reality is that, on average, up to 5-15% of hard drives actually fail within one year ([Backblaze, 2014](#)), ([Pinheiro et al, 2007](#)).
- Because failures and errors are relatively rare events, reliability statistics from vendors are typically based on models and simulations and not from long-term observations of what actually happens in practice. For example, if a manufacturer says that the shelf life of media is 30 years then it's not because they have actually tested media over that time period. Likewise, if a vendor estimates the MTTL is 1,000 years then they clearly haven't built a system and tested it for anywhere near that length of time. Therefore, statistics should be interpreted as best estimates from vendors of how a system might behave in practice - but it may not actually turn out that way. For example, field studies have suggested that manufacturer estimates of reliability can be over optimistic ([Jiang et al, 2008](#)) .
- The likelihood of data loss increases dramatically when correlations are taken into account. Correlations are where parts of a system, or different copies of the digital material, can't be considered as independent. If there is a problem with one part of the system or copy of the digital material then there is likely to be a problem with another part or copy. Examples include a manufacturing fault affecting all the hard drives in a storage server, software or firmware bugs systemically corrupting digital material, failure by an organisation to regularly test its backups, or failure to isolate or decouple storage systems so that if one copy of the digital material is accidentally deleted then all the other copies don't get deleted too. These correlation factors can be far more significant than the specific failure modes covered by reliability statistics.

These findings and observations result in the following recommendations:

- Plan for failures to happen in IT storage solutions no matter how cleverly designed by the manufacturer. Failure rates in practice may well be higher than manufacturer statistics suggest.
- Data loss can be caused by failure to put in place proper processes and procedures around the use of IT storage as well as from the storage technology itself. Proper risk assessment is the way to identify and manage these problems.
- The best strategy remains to create multiple independent copies of digital material in different locations and to store them using different technologies where possible. This should include a process of actively and regularly checking data integrity of all the copies so problems can be detected no matter why and where they might occur. In this way, risks are both minimised and spread, and reliance isn't placed on any particular storage technology or service being completely error free.

Multi-copy storage strategies

Digital storage technologies present several risks to long-term preservation of digital objects. These risks can be reduced by using a digital storage strategy that involves one or more storage systems and at least two copies of the data.

Good practice is for a storage strategy to have the following characteristics:

- (a) multiple independent copies exist of the digital materials
- (b) these copies are geographically separated into different locations
- (c) the copies use different storage technologies
- (d) the copies use a combination of online and offline storage techniques
- (e) storage is actively monitored to ensure any problems are detected and corrected quickly.

A digital storage strategy can be implemented in a staged way, starting with a basic level of protection and access to digital content and moving on towards a more automated and scalable approach that gives a higher level of data safety and security.

Risks to digital content come from a range of sources and a digital storage strategy helps balance the cost of digital storage with the reduction of those risks. Example risks to consider include fire, flood, failure to instigate or follow proper processes or procedure, malicious attack, media degradation, and obsolescence of storage systems and technologies. The principal risks and means of addressing or mitigating them are often addressed in an organisation's business continuity planning (see [Risk and change management](#)).

It is important to realise that many examples of content loss are not necessarily due to technical faults with storage technology (although it is important to recognise that these do happen), but can come from human error, lack of budget or planning of storage migrations, or a failure to regularly check and correct failures that might occur.

In a world that is increasingly using networked systems and technologies for digital storage, there is a role for an offline copy of digital materials. This can provide a 'fire break' against problems with online systems that can automatically propagate between locations, e.g. deletion of a file in one location that automatically deletes a mirrored copy at another site.

Making more than one copy of the digital materials is fundamental to achieving a basic level of data safety. Using different types of storage for each copy helps spread the risk and ensure that a problem with one technology doesn't affect the others. The way each copy is stored can be adjusted to achieve an acceptable overall level of cost, risk and complexity. For example, one copy might be held using an online storage server for fast access and one copy might be on data tape in deep archive for low cost and relatively high safety.

This Handbook follows the National Digital Stewardship Alliance (NDSA) preservation levels ([NDSA, 2013](#)) below in recommending four levels at which digital preservation can be supported through storage and geographic redundancy. We make the additional recommendation of using a combination of online and offline copies to achieve a good combination of data access and data safety:

Level	Approach	Risks addressed and benefits achieved
1	<ul style="list-style-type: none"> Two complete copies of the digital materials that are not co-located. One copy should be offline. For digital materials on heterogeneous media (optical disks, hard drives, etc.) get the content off the medium and into your storage system. 	<ul style="list-style-type: none"> Basic ability to recover from a range of issues including storage system failure. Loss or damage to one copy can be recovered using the other copy. Digital materials easier to manage when in a single storage system.
2	<ul style="list-style-type: none"> At least three complete copies. At least one copy in a different geographic location. Document your storage system(s) and storage media and what you need to use them. 	<ul style="list-style-type: none"> As above plus protection from natural disasters and other major events. Good level of access and digital materials safety. Staff have clear policies and procedures to follow so are more efficient, costs are lowered, and staff changes can be managed.
3	<ul style="list-style-type: none"> At least one copy in a geographic location with a different disaster threat. Obsolescence monitoring and migration process for your storage system(s) and media. 	<ul style="list-style-type: none"> As above plus protection from the longer-term risks associated with technical obsolescence. Continual access to content is possible even during migrations and disasters.
4	<ul style="list-style-type: none"> At least three copies in geographic locations with different disaster threats. Have a comprehensive plan in place that will keep files and metadata on currently accessible media or system. 	<ul style="list-style-type: none"> As above with full range of risks addressed including accidental loss and malicious attack, vendor lock-in, and budget instabilities. Content has high availability, costs are predictable and manageable, there is the ability to achieve trusted repository certification.

Managing storage system obsolescence and risks

The use of storage technologies and solutions needs careful planning and management to be an effective approach to supporting digital preservation. If done properly, the result can be very good levels of data safety, rapid access to content when needed, and costs that are both low and predictable.

IT storage technologies can fail or cause data corruption and the lifetime of media and systems is typically short, for example 3-5 years, which means solutions become obsolete quickly and migration is needed to avoid digital materials becoming at risk. Migration in this context means moving data

off an old storage system and onto a new storage system. The digital material itself does not change but the storage solution does. An IT department or storage service provider will think of migration at the storage level. This is in contrast to file format migration where the file format will change, but the way that the files are stored doesn't change.

Resources



NDSA Levels of Preservation

<http://www.digitalpreservation.gov/ndsa/activities/levels.html>

The National Digital Stewardship Alliance (NDSA) "Levels of Digital Preservation" are a tiered set of recommendations for how organizations should begin to build or enhance their digital preservation activities. It is intended to be a relatively easy-to-use set of guidelines useful not only for those just beginning to think about preserving their digital assets, but also for institutions planning the next steps in enhancing their existing digital preservation systems and workflows. It is not designed to assess the robustness of digital preservation programs as a whole since it does not cover such things as policies, staffing, or organizational support.



These are some of the more notable digital preservation storage systems and storage system/service providers. There are a wide-range of commodity IT storage vendors, as well as specialist digital preservation service providers that can provide onsite or cloud storage (see also [Cloud services](#)). These specialists typically may support other preservation functions in addition to storage.

Arkivum

<http://arkivum.com>

Digital Preservation Network

<http://www.dpn.org>

DSpace

<http://www.dspace.org>

ePrints

<http://www.eprints.org>

Fedora

<http://fedorarepository.org>

iRods

<http://irods.org>

LOCKSS

<http://www.lockss.org>

OCLC Digital Archive

<http://www.oclc.org/digital-archive.en.html>

Portico

<http://www.portico.org/digital-preservation/>

Preservica

<http://preservica.com>

Rosetta

<http://www.exlibrisgroup.com/category/RosettaOverview>

Community Owned digital Preservation Tool Registry COPTR

http://coptr.digipres.org/Main_Page

Although focussing principally on tools the COPTR registry also covers a range of storage systems and services. It acts primarily as a finding and evaluation tool to help practitioners find the tools they need to preserve digital data. COPTR captures basic, factual details about a tool, what it does, how to find more information (relevant URLs) and references to user experiences with the tool.



DSHR's Blog

<http://blog.dshr.org>

David Rosenthal is a computer scientist and chief scientist for the LOCKSS project. His blog frequently covers computer storage development and trends and implications for digital preservation.

Case studies



The National Archives case study: Bodleian Library, University of Oxford

<http://www.nationalarchives.gov.uk/documents/archives/case-study-oxford.pdf>

This case study covers the Bodleian Library and the University of Oxford, and the provision of a "private cloud" local infrastructure for its digital collections including digitised books, images and multimedia, research data, and catalogues. It explains the organisational context, the nature of its digital preservation requirements and approaches, its storage services, technical infrastructure, and

the business case and funding. It concludes with the key lessons they have learnt and future plans. January 2015 (4 pages).

The National Archives case study: Parliamentary Archives

<http://www.nationalarchives.gov.uk/documents/archives/case-study-parliament.pdf>

This case study covers the Parliamentary Archives. It is an example of an archive using a hybrid set of storage solutions (part-public cloud and part-locally installed) for digital preservation as the archive has a locally installed preservation system (Preservica Enterprise Edition) which is integrated with cloud and local storage and is storing sensitive material locally, not in the cloud. January 2015 (4 pages).

The National Archives case study: Tate Gallery

<http://www.nationalarchives.gov.uk/documents/archives/case-study-tate-gallery.pdf>

This case study discusses the experience of developing a shared digital archive for the Tate's four physical locations powered by a commercial storage system from Arkivum. It explains the organisational context, the nature of their digital preservation requirements and approaches, and their rationale for selecting Arkivum's on-premise solution, "Arkivum/OnSite" in preference to any cloud-based offerings. It concludes with the key lessons learned, and discusses plans for future development. January 2015 (4 pages).

Use Cases: Storage Solutions for Digital Preservation

<http://www.alliancepermanentaccess.org/index.php/aparsen/webinars/#09>

A set of powerpoint presentations covering use cases and survey results from an APARSEN event on 14 April 2014. The uses cases for the National Library of the Netherlands (KB) and RSS Data Farm (European Space Agency) are particularly relevant.

References

Backblaze, 2014. Hard Drive Reliability Update – Sep 2014. *Backblaze*. [blog] Available:

<https://www.backblaze.com/blog/hard-drive-reliability-update-september-2014/>

Elerath, J., 2009. Hard-Disk Drives: The Good, the Bad, and the Ugly. *Communications of the ACM*. 52 (6), 38-45. Available: doi:10.1145/1516046.1516059.

<http://cacm.acm.org/magazines/2009/6/28493-hard-disk-drives-the-good-the-bad-and-the-ugly/fulltext>

Greenan, K.M., Plank, J.S. & Wylie, J.J., 2010. Mean time to meaningless: MTDL, Markov models, and storage system reliability. *Proceedings of the 2nd USENIX conference on Hot topics in storage and file systems*. Available:

https://www.usenix.org/legacy/event/hotstorage10/tech/full_papers/Greenan.pdf

Jiang, W. et al., 2008. Are Disks the Dominant Contributor for Storage Failures? A Comprehensive Study of Storage Subsystem Failure Characteristics. *Proceedings of the 6th USENIX Conference on File and Storage Technologies (FAST '08)*. Available:

<http://www.usenix.org/events/fast08/tech/jiang.html>

NDSA, 2013. *The NDSA Levels of Digital Preservation: An Explanation and Uses, version 1 2013*.

National Digital Stewardship Alliance. Available:

http://www.digitalpreservation.gov/ndsa/working_groups/documents/NDSA_Levels_Archiving_2013.pdf

Pinheiro, P., Weber, W-D. & Barroso, L.A., 2007. Failure Trends in a Large Disk Drive Population. *Proceedings of the 5th USENIX Conference on File and Storage Technologies (FAST' 07)*. Available: http://static.googleusercontent.com/media/research.google.com/en//archive/disk_failures.pdf

Rosenthal, D.S.H., 2010. Bit Preservation: A Solved Problem? *The International Journal of Digital Curation*. 5 (1) Stanford University Libraries, CA. Available: <http://www.ijdc.net/index.php/ijdc/article/view/151/224>

Legacy Media



Illustration by Jørgen Stamp digitalbevaring.dk CC BY 2.5 Denmark

Introduction

Many organisations will have large amounts of data stored on legacy media, such as magnetic and optical media, and data will continue to be received on old carriers. Ultimately, the best long-term strategy for the preservation of the data will be migration to file-based storage and active management thereafter (see [Storage](#) section). Often the original media will continue to be preserved alongside this, so it will be necessary to understand their preservation and storage requirements. For organisations with large collections of legacy media, understanding the risks facing each media type will also help with prioritising collections for migration and application of digital forensics tools and methods will also be helpful (see [Digital forensics](#) section).

For the preservation of magnetic and optical media, two aspects need to be considered - the media itself and the hardware and software needed to interpret it. In some cases the second aspect will be the most challenging. As the popularity of a media format declines, the manufacture of hardware ceases and becomes more difficult to procure and maintain.

Preserving legacy media

In most cases, the simplest way to mitigate risks with storage media is to transfer all content into a managed storage system. This means that the content can be managed without reference to the original storage medium. This would probably be adequate for the vast majority of digital content requiring preservation. However, there may be a few instances where it is necessary to retain the original media carrier in some way. In some cases, the storage medium could simply be retained as an artifact, with no expectation of long-term access, e.g. where it forms part of a hybrid collection or has some kind of value by association. (e.g. part of the collections of a prominent author). However, where continued access to the content is required, careful thought needs to be given to how it could be accessed in the future.

One thing that we do know from experience is that digital storage media types change frequently over time. For example, the previous version of this handbook contained an overview of magnetic and optical storage media and provided estimates of the lifetimes of selected storage media types that were popular in the mid-1990's (a digital preservation handbook written in previous decades would presumably have included assessments of punched cards and paper tape). Given current trends in storage technology, it is perhaps better now to provide a framework that supports the ongoing evaluation of storage media, which might now include flash memory sticks or external hard drives. One such framework has been provided by The National Archives ([Brown, 2008](#)). This uses a scorecard approach, measuring selected storage media against six criteria:

- longevity (e.g., proven operational lifetimes)
- capacity
- viability (e.g., in terms of retaining evidential integrity)
- obsolescence
- cost
- susceptibility (e.g., to physical damage and to different environmental conditions).

In practice, however, these kinds of assessment can only get you so far. There is a growing body of evidence that suggests that variation in manufacturing quality also plays a major role in media longevity ([Harvey, 2011](#)). That is why, in the end, digital preservation normally depends upon the transfer of content from media into a managed storage environment.

Resources



Selecting storage media for long-term preservation, TNA Digital Preservation Guidance Note 2: August 2008

<https://www.nationalarchives.gov.uk/documents/selecting-storage-media.pdf>

This document is one of a series of guidance notes produced by The National Archives, giving general advice on issues relating to the preservation and management of electronic records. It is intended for use by anyone involved in the creation of electronic. It provides information for the creators and

managers of electronic records about the selection of physical storage media in the context of long-term preservation. Note guidance is as of August 2008. (7 pages).

Care, Handling and Storage of Removable media, TNA Digital Preservation Guidance Note 3: August 2008

<http://www.nationalarchives.gov.uk/documents/information-management/removable-media-care.pdf>

This document is one of a series of guidance notes produced by The National Archives, giving general advice on issues relating to the preservation and management of electronic records. It provides advice on the care, handling and storage of removable storage media. Note guidance is as of August 2008. (10 pages).

You've Got to Walk Before You Can Run: First Steps for Managing Born-Digital Content Received on Physical Media

<http://www.oclc.org/content/dam/research/publications/library/2012/2012-06.pdf>

A step by step guide about getting digital born material off of various physical media. It focuses on identifying and stabilizing your holdings so that you'll be in a position to take additional steps as resources, expertise, and time permit. The POWRR project document [Resources for Technical Steps](#) (3 pages) adds additional resources for some of the steps. (7 pages).



Kryoflux: Commercial tool for reading floppy disks

<http://www.kryoflux.com/>

KryoFlux is a USB-based device designed specifically to acquire reliable low-level reads suitable for software preservation. This is the official hardware developed by The Software Preservation Society,



Digital Preservation Management: Chamber of Horrors

<http://dpworkshop.org/dpm-eng/oldmedia/disks.html>

Some examples of obsolete and endangered disks.

Lost Formats

<http://www.experimentaljetset.nl/archive/lostformats>

Web page from the Lost Formats Preservation Society with a very nice overview of silhouettes of the shapes to allow quick identification and key brief history and features such as dimensions and storage capacity. All silhouettes shown as same size rather than to scale. Last major update appears to be c.2008 but content is still valuable for all but the most recent formats.

Museum Of Obsolete Media

<http://www.obsoletemedia.org/category/format/>

Great resource covering a very wide-range of obsolete audio, video, data, and film storage media. You can browse the categories or the Gallery and Timeline. Particularly good if you know what you are looking for and derived mostly from the relevant Wikipedia entries.

Case studies



A Fistful of Floppies: Digital Preservation in Action

https://ischool.uw.edu/sites/default/files/capstone/posters/JStanley_Capstone_Landscape.pdf

The University of Washington Library system currently holds a small collection of electronic thesis and dissertation (ETD) accompanying materials from the late 1980's to 2011 on floppy disks and CD-Rs. These materials will soon reach or have already exceeded the limit of their expected lifespans. This 2015 project looked at the digital preservation possibilities for this collection of materials using digital forensics as a model.(1 page).

Enford, D., et al 2008, Media Matters: developing processes for preserving digital objects on physical carriers at the National Library of Australia, Papers from 74th IFLA General Conference and Council

<http://archive.ifla.org/IV/ifla74/papers/084-Webb-en.pdf>

The National Library of Australia had a relatively small but important collection of digital materials on physical carriers, including both published materials and unpublished manuscripts in digital form. The Digital Preservation Workflow Project aimed to produce a semi-automated, scalable process for transferring data from physical carriers to preservation digital mass storage, helping to mitigate the major risks associated with the physical carriers. (17 pages).

Digital Preservation Planning Case Study

http://www.dpconline.org/component/docman/doc_download/863-2013-may-getting-started-london-planning-case-study-ed-fay

Presentation on getting started with digital preservation planning, including scoping, risk assessing and prioritising your collection (including legacy media examples), and staff roles and responsibilities. 2013 (20 pages).

References

Brown, A., 2008. Selecting storage media for long-term preservation. *TNA Digital Preservation Guidance Note 2: August 2008*. Available:

<https://www.nationalarchives.gov.uk/documents/selecting-storage-media.pdf>

Harvey, R., 2011. *Preserving Digital Materials 2nd edition*. De Gruyter Saur.

Preservation Planning

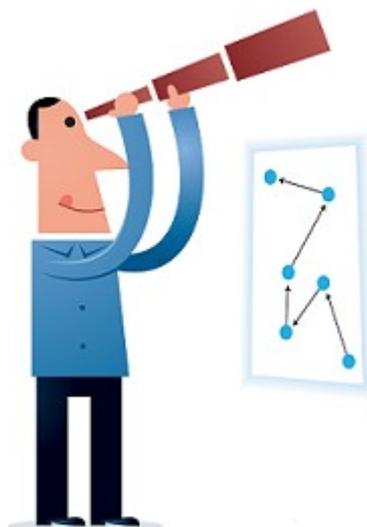


Illustration by Jørgen Stamp digitalbevaring.dk CC BY 2.5 Denmark

What is preservation planning?

Preservation planning is the function within a digital repository for monitoring changes that may impact on the sustainability of, or access to, the digital material that the repository holds. It should be proactive: both current and forward-looking in terms of acquisitions and trends. Changes might occur within the repository, within the organisation in which the repository resides, or external to the repository and organisation themselves. Changes might be monitored in the following areas:

- Technology watch
 - Packaging
 - Storage
 - Formats
 - Tools
 - Environment
 - access mechanisms
- Designated communities
 - needs and expectations of users
 - needs and expectations of producers
 - emerging tools for machine to machine access
 - formal feedback from users and producers

The concept of preservation planning is defined within the functional model of the [OAIS](#) standard ([CCSDS, 2012](#)). This section focuses primarily on the Monitoring components within the OAIS definition. The 'Monitor Technology' and 'Monitor Designated Community' functions of OAIS provide surveys that inform preservation planning activities. These alert the repository about changes in the external environment and risks that could impact on its ability to preserve and maintain access to the information in its custody, such as innovations in storage and access technologies, or shifts in the scope or expectations of the Designated Community (see [Lavoie, 2015,13](#)). Preservation planning then develops recommendations for updating the repository's policies and procedures to

accommodate these changes. The Preservation planning function represents the OAIS's safeguard against a constantly evolving user and technology environment. It detects changes or risks that impact the repository's ability to meet its responsibilities, designs strategies for addressing them, and assists in the implementation of these strategies within the archival system.

What is the purpose of preservation planning?

Identifying triggers for taking action to preserve digital materials

Where change has been identified, a risk assessment process can be used to analyse and identify the change that represents a significant risk to the digital material in the repository. Risks can then be addressed and hopefully mitigated following a preservation planning exercise to decide on appropriate preservation action. In this case, the monitoring or technology watch process is identifying trigger points for further analysis, preservation planning and, where relevant, action to preserve digital materials.

Building a knowledge base to inform preservation activities

The process of monitoring internal and external factors as part of a preservation planning activity can inform the knowledge base of an organisation, and in doing so improve its ability to perform digital preservation activities effectively. For example the "knowledge base" of an organisation might be augmented with information about the capabilities of a new software tool, or the obsolescence and unavailability of an existing tool. In some cases this process might be best performed individually or within an organisation, but alternatively might be more usefully performed in a collaborative manner. The vast depth and breadth of knowledge required for digital preservation naturally favours a collaborative approach, whereby particular organisations are able specialise in a particular area and contribute that knowledge to an open or shared knowledge base.

Implementations of a preservation planning service

The degree to which technology watch will be necessary will vary according to the degree of uniformity or control over formats and media that can be exercised by the institution. Those with little control over media and formats received and a high degree of diversity in their holdings will find this function essential. For most other institutions the IS strategy should seek to develop corporate standards so that everybody uses the same software and versions and is migrated to new versions as the products develop.

Failure to implement an effective technology watch or IS strategy incorporating this will risk potential loss of access to digital holdings and higher costs. It may be possible for example to re-establish access through digital forensics (see [Digital forensics](#)) but this may be expensive compared to pre-emptive strategies.

A retrospective survey of digital holdings (see [Getting started](#)) and a risk assessment and action plan (see [Risk and change management](#)) may be a necessary first step for many institutions, prior to implementing a technology watch.

Good preservation metadata in a computerised catalogue identifying the storage medium, the necessary hardware, operating system and software will enable a technology watch strategy (see [Metadata and documentation](#)).

Integrated preservation systems, and individual tools and registries can also support this function (see [Technical solutions and tools](#)).

Resources

Some of the core preservation watch activities are generic and therefore ready made for collaboration while others are highly localised and not easily shared.



DPC Technology Watch Report Series

<http://www.dpconline.org/advice/technology-watch-reports>

These reports provide an advanced introduction to specific issues for those charged with establishing or running services for long term preservation and access. They are updated and new reports added periodically.

Scout – a preservation watch system, OPF blog post 16th Dec 2013

<http://openpreservation.org/blog/2013/12/16/scout-preservation-watch-system/>

The SCAPE Project designed a demonstrator for an automated preservation watch service, called SCOUT. SCOUT was described by its developers as providing "...an ontological knowledge base to centralize all necessary information to detect preservation risks and opportunities. It uses plugins to allow easy integration of new sources of information, as file format registries, tools for characterization, migration and quality assurance, policies, human knowledge and others."

Assessing file format risks: searching for Bigfoot? OPF Blog post 29th Oct 2014

<http://openpreservation.org/blog/2013/09/30/assessing-file-format-risks-searching-bigfoot/>

This detailed blog post raises concerns about challenges with automating preservation watch.

Barbara Sierman, Paul Wheatley 2010 Evaluation of Preservation Planning within OAIS, based on the Planets Functional Model Planets Deliverable no. PP7-D6.1

http://www.planets-project.eu/docs/reports/Planets_PP7-D6_EvaluationOfPPWithinOAIS.pdf

The Planets Project realised various aspects of the concepts defined within the OAIS Preservation Planning function, and performed an evaluation of OAIS based on these practical experiences. 2010 (34 pages).



Community Owned digital Preservation Tool Registry COPTR

http://coptr.digipres.org/Main_Page

COPTR describes tools useful for long term digital preservation and acts primarily as a finding and evaluation tool to help practitioners find the tools they need to preserve digital data. COPTR aims to collate the knowledge of the digital preservation community on preservation tools in one place. It was initially populated with data from registries run by the COPTR partner organisations, including

those maintained by the Digital Curation Centre, the Digital Curation Exchange, National Digital Stewardship Alliance, the Open Preservation Foundation, and Preserving digital Objects With Restricted Resources project (POWRR). COPTR captures basic, factual details about a tool, what it does, how to find more information (relevant URLs) and references to user experiences with the tool. The scope is a broad interpretation of the term "digital preservation". In other words, if a tool is useful in performing a digital preservation function such as those described in the OAIS model or the DCC lifecycle model, then it's within scope of this registry.

Case studies



OCLC Research Report - Preservation Health Check: Monitoring Threats to Digital Repository Content

<http://www.oclc.org/research/themes/research-collections/phc.html>

The OCLC Research Preservation Health Check activity was initiated by Open Planets Foundation. The Pilot used a sample of preservation metadata provided by the Bibliothèque Nationale de France. The report presents the preliminary findings of Phase 1 of the Pilot and suggests that there is an opportunity to use PREMIS preservation metadata as an evidence base to support a threat assessment exercise based on the Simple Property-Oriented Threat (SPOT) model.

Digital Preservation Planning Case Study

http://www.dpconline.org/component/docman/doc_download/863-2013-may-getting-started-london-planning-case-study-ed-fay

Presentation on getting started with digital preservation planning, including scoping, risk assessing and prioritising your collection (including legacy media examples), and staff roles and responsibilities. 2013 (20 pages).

References

Consultative Committee for Space Data Systems, 2012. *Reference model for an open archival information system (OAIS): Recommended practice* (CCSDS 650.0-M-2: Magenta Book), CCSDS, Washington, DC. Available: <http://public.ccsds.org/publications/archive/650x0m2.pdf>

(Note this is a free to download version of ISO 14721:2012, Space Data and Information Transfer Systems – Open Archival Information System (OAIS) – Reference Model, 2nd edn).

Lavoie, B., 2014. *The Open Archival Information System (OAIS) Reference Model: Introductory Guide (2nd Edition)* DPC Technology Watch Report 14-02 October 2014. Available: <http://dx.doi.org/10.7207/TWR14-02>

Preservation Action



Illustration by Jørgen Stamp digitalbevaring.dk CC BY 2.5 Denmark

Introduction

We know by now that digital preservation is comprised of a series of challenges emanating from organisational, resourcing, managerial, cultural, and technical issues. This section of the Handbook will focus specifically on actions that can be taken to help mitigate the technical challenges of preserving digital materials over time.

Technological obsolescence of formats

Technological obsolescence has long been considered a significant challenge of long term digital preservation. However in recent years studies have suggested that format obsolescence isn't always as prevalent as previously feared ([Rosenthal 2015a](#), [Jackson 2012](#)). It is one issue that must be recognised and countered if digital materials are to survive over generations of technological change but it is certainly not the only challenge. Many established file formats are still with us, still supported, and still usable. It is quite likely that the majority of file formats you deal with will be commonly understood and well supported rather than obsolete.

A simple definition of obsolescence is the process of becoming outdated or no longer used. When talking about technological obsolescence, we refer for example to 'this Wordperfect 3.1 software is obsolete' or 'this BBC Micro computer is obsolete'. The exact moment at which obsolescence occurs can be difficult to pinpoint, particularly for materials that have only recently become obsolete. For example, just because the original application (e.g. MS Word) no longer supports a given format, it doesn't mean no other software that can read the format is unavailable. Similarly one institution may continue to use and maintain a piece of legacy software long after others have upgraded to new versions. It is perhaps therefore more useful to talk about 'institutional obsolescence', namely that the technology in question is no longer in use or easily accessed by a particular institution.

Obsolescence is an issue because all files have their own hardware and software dependencies. This was particularly the case in the early days of computing.

Change becomes an issue when it compromises the meaning of the content or its interpretation by a user. A core goal of digital preservation actions is to preserve the integrity and authenticity of the material being preserved, despite these generational changes in computing technology. In the next section we will discuss some common strategies to help minimise these changes.

Preservation strategies

In this section we review the technical strategies that can be employed to preserve digital information. After a flurry of activity in the late 1990s there has been relatively little progress in finding new strategies, though there has been significant research and development into varying implementation options and supporting technologies such as quality assurance, digital forensics (see [Digital forensics](#)), and technical representation information registries (see [Technical solutions and tools](#) in the Handbook). The techniques we will cover here are:

- Format Migration
- Emulation
- Computer Museums

Format migration

Format Migration is one of the most widely utilised preservation strategies employed and most digital preservation systems contain functionality or system data that assumes a migration solution. Format migration is different from storage media migration. It involves transferring or transforming (i.e. migrating) data from an ageing/obsolete format to a new format, possibly using new applications systems at each stage to interpret information. Moving from one version of a format standard to a later standard is a version of this method; for example moving from MS Word Version 6 (from 1993) to MS Word for Windows 2010. For frameworks and tools that are helpful for evaluating technical obsolescence of file formats see [File formats and standards](#).

Format migration, like any intervention that has the potential to change the structure and content of data, can introduce errors and loss of information. Therefore, it is important to define metrics to measure possible loss of information and use these to do tests on the correctness and quality of format migration.

Recent work touching on quality assurance and digital preservation actions includes the work of the [AQUA](#), [SPRUCE](#), and [SCAPE](#) projects. To measure error rates, it is necessary to determine some very specific metrics. You might need to define what you count as an error and whether you weight some errors as being more important than others. This depends on the context/content of the record and what characteristics of the material are deemed 'significant' to preserve, as well as the migration tools and successive formats used in any migration pathway.

Some practical issues involved in this process include when to migrate – is it better to migrate from generation to generation, or should some generations be skipped? You will need to keep a record of all transformations, their results and to document detected losses of information so as to maintain evidence of authenticity and authority. PREMIS can be a useful tool for this - see the Handbook section on [Metadata and documentation](#) for more information about this standard. It is good practice always to retain the original file format as deposited to return to if required.

Emulation

Emulation offers an alternative solution to migration that allows archives to preserve and deliver access to users directly from original files. This technique attempts to preserve the original behaviours and the look and feel of applications, as well as informational content. It is based on the view that only the original programme is the authority on the format and this is particularly useful for complex objects with multiple interdependencies, such as games or interactive apps.

An emulator, as the name implies, is a programme that runs on a current computer architecture but provides the same facilities and behaviour as an earlier one. This approach has been endorsed by a number of heritage organisations, often in collaboration with technical experts and in recent years there has been some notable success in implementing emulation solutions for cultural heritage (see [Resources](#) below) . However some significant challenges remain, not least there are often rights issues associated with software licensing that need to be resolved ([Rosenthal 2015b](#)).

A particular benefit of emulation is that a single solution can be deployed to provide access to a large number of objects, so long as all those objects require delivery on the same operating system or hardware stack. Use of legacy computing equipment may however prove difficult for users, though they will almost certainly be accessing an 'authentic' representation of the records. Of course emulators have to be built and maintained, requiring a pool of expertise to be available and this cannot always be assumed. New emulators will be needed as computer architectures become obsolete, and both of these present costs and resource needs.

Computer museums

This methodology proposes the keeping of computers and their systems software (operating systems, drivers, etc.) as well as the data and applications programmes. Effort must be expended to keep all platforms in good order, and to retain all the knowledge necessary to maintain and use the machines and their programmes. The idea relies on having a source of spare parts too, but these will dwindle, as will pools of expertise. Hence this strategy tends to be an interim measure rather than a long-term solution. Some formal museums do exist, such as the [Computer History Museum](#) in California and the [Centre for Computing History](#) in Cambridge. These typically maintain machines in working order though do not provide preservation services. See also the [Legacy media](#) section of the Handbook for further information on historic file formats and media.

Implementation

The [DPC Technology Watch Reports](#) are a particularly useful guide to most common genres and file formats (including email, social media, Audio-Visual, eBooks, e-Journals, GIS, CAD, web archiving etc.) and show which strategies tend to be used most commonly in each of these areas. Tools to assist with implementation of preservation strategies are discussed in the [Technical solutions and tools](#) area of the Handbook particularly in [File formats and standards](#).

Resources



DPC Technology Watch Report series

<http://www.dpconline.org/publications/technology-watch-reports>

The DPC Technology Watch Report series is intended as an advanced introduction to specific issues for those charged with establishing or running services for long term access. They identify and track developments in IT, standards and tools which are critical to digital preservation activities. They are commissioned by experts on these developments and are thoroughly scrutinised by peers before being released.

Emulation & Virtualization as Preservation Strategies

https://mellon.org/media/filer_public/0c/3e/0c3eee7d-4166-4ba6-a767-6b42e6a1c2a7/rosenthal-emulation-2015.pdf

This 2015 report on Emulation and Virtualization as Preservation Strategies by David Rosenthal was funded by the Mellon Foundation, the Sloan Foundation and IMLS. It concludes recent developments in emulation frameworks make it possible to deliver emulations to readers via the Web in ways that make them appear as normal components of Web pages. This removes what was the major barrier to deployment of emulation as a preservation strategy. Barriers remain, the two most important are that the tools for creating preserved system images are inadequate, and that the legal basis for delivering emulations is unclear, and where it is clear it is highly restrictive. Both of these raise the cost of building and providing access to a substantial, well-curated collection of emulated digital artefacts beyond reach. If these barriers can be addressed, emulation will play a much greater role in digital preservation in the coming years. (37 pages).

Systematic planning for digital preservation: evaluating potential strategies and building preservation plans

<http://www.ifs.tuwien.ac.at/~strodl/paper/becker-ijdl2009.pdf>

This article published in 2009 describes a systematic approach for evaluating potential alternatives for preservation actions and building thoroughly defined, accountable preservation plans for keeping digital content alive over time. The work was undertaken as part of the European Union-funded PLANETS project. (25 pages).

File format conversion

<http://www.nationalarchives.gov.uk/documents/information-management/format-conversion.pdf>

Format conversion may help you maintain access and use of your information and mitigate risks that arise from obsolescence. This 2011 guidance from The National Archives gives you the steps you should go through in performing a file format conversion process. (29 pages).

What organizations are preserving software

<http://qanda.digipres.org/1068/what-organizations-are-preserving-software>

This post and responses from August 2015 on the Digital Preservation Q&A site provides a useful list and links for institutions preserving software for emulation strategies.

SCAPE Project Final best practice guidelines and recommendations

http://scape-project.eu/wp-content/uploads/2014/02/SCAPE_D20.6_KB_V1.0.pdf

This SCAPE project report published in 2014 covers three major areas: implementation of large-scale migration as a preservation strategy. Other areas are preservation of research data; and Bit preservation. (127 pages).

Case studies



The Internet Arcade

<https://archive.org/details/internetarcade>

The Internet Arcade is a web-based library of arcade (coin-operated) video games from the 1970s through to the 1990s from the Internet Archive, implemented using an in-browser emulation solution to provide access to the collection.

Interject

<http://www.webarchive.org.uk/interject/>

This prototype from the British Library demonstrates how a mixture of preservation actions can be smoothly integrated into the search infrastructure of the UK Web Archive, by acting as an 'access helper' for end users. This web page picks a few specific examples of difficult or interesting cases, and allows you to inspect what the system knows about those formats. You can also view transformed versions of those resources (combining format conversion and emulation techniques).

Rhizome

<http://rhizome.org/editorial/2015/apr/17/theresa-duncan-cd-roms-are-now-playable-online/>

In the 1990s, Theresa Duncan and collaborators made three videogames that exemplified interactive storytelling at its very best. Two decades later, the works (like most CD-ROMs) have fallen into obscurity. This online exhibition, co-presented by Rhizome and the New Museum brings them back, making them playable online via emulation.

Assessing Migration Risk for Scientific Data Formats

<http://www.ijdc.net/index.php/ijdc/article/view/202/271>

This paper explore a simple hypothesis – that, where migration paths exist, the majority of scientific data files can be safely migrated leaving only a few that must be handled more carefully – in the context of several scientific data formats that are or were widely used. The approach is to gather information about potential migration mismatches and, using custom tools, evaluate a large collection of data files for the incidence of these risks. The results support the initial hypothesis, though with some caveats.

Portico - Preservation Step-by-Step

<http://www.portico.org/digital-preservation/services/preservation-approach/preservation-step-by-step>

A useful step by step guide to the preservation planning and migration strategies employed by Portico The preservation plan may include an initial migration of the packaging or files in specific formats (for example, Portico migrates publisher specific e-journal article XML to the NLM archival standard).

Trash to treasure: Retro computer, software collection helps National Library access digital pieces

<http://www.abc.net.au/news/2015-06-20/collecting-retro-computer-technology-to-save-digital-treasures/6560494>

The National Library of Australia made public its own efforts to develop a collection of legacy computing hardware and software. It uses it to support data recovery and then implements other preservation strategies and does not rely on the computer museum for long-term preservation.

References

David Rosenthal, 2015a. *"The Prostate Cancer of Preservation" Re-examined*. Available: <http://blog.dshr.org/2015/09/the-prostate-cancer-of-preservation-re.html>

David Rosenthal, 2015b. *Emulation & Virtualization as Preservation Strategies*. Available: https://mellon.org/media/filer_public/0c/3e/0c3eee7d-4166-4ba6-a767-6b42e6a1c2a7/rosenthal-emulation-2015.pdf

Andrew N. Jackson, 2012. *Formats over Time: Exploring UK Web History*,. Available: <http://arxiv.org/abs/1210.1714>

Access



Illustration by Jørgen Stamp digitalbevaring.dk CC BY 2.5 Denmark

Introduction

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 both current and future generations. Preserving access to digital objects is the key objective of digital preservation programmes but requires more active management throughout the lifecycle of the resource before it can be assured. It is, therefore, essential to consider issues important to access provision from the beginning of the preservation process, ideally as early as the acquisition phase. This is represented within the [Decision Tree for Selection of Digital Material](#) for Long-Term

Retention included within the [Acquisition and appraisal](#) section of this handbook. With this in mind this section aims to identify the main issues that must be considered, the decisions that should be made when planning for access provision and how these may impact on preservation more generally.

Understanding users

Understanding potential users is essential when planning for the provision of access to digital objects as well as being a key consideration of broader digital preservation activities. The importance of such work is perhaps most evident in the focus on 'Designated Communities' within the Open Archival Information System Reference Model. Knowledge gained about these potential users will inform decisions made throughout the lifecycle but will likely hold most weight when choosing suitable access delivery solutions, balanced with resource and technological considerations. It is important to approach the identification of user communities and their needs systematically and objectively. In short, understanding what users want to do and what functionality can be provided by the repository.

The methodology used for the gathering of this information will vary depending on the organisational context. Potential options and tools may include the following:

- Analysis of current usage (access requests for both physical and digital objects, website statistics etc.)
- Surveys
- Focus groups
- Interviews
- Use cases
- Task analysis

When carrying out user analysis it is important to consider both existing users and non-users. Although interaction with non-users is inherently more difficult this can be a useful process towards understanding current barriers to use as well as identifying potential new market sectors.

Once collected this information should be used to inform decisions that are made in relation to the implementation of access delivery solutions. It is also important to continue to monitor the development of user communities and this should be incorporated in the standard [Preservation planning](#) activities within your organisation.

Access formats

A key consideration when planning for access is the format in which the digital objects will be delivered to the users. 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, so it may be desirable or even necessary, to deliver an access copy of the digital object to the user in a different format from that held within the preservation system's storage. Indeed, an organisation may wish to offer different 'flavours' of format depending on the needs of the particular user or community in question. When selecting formats for access there are several questions an organisation will need to consider, these may include the following:

- What is the mostly commonly used/widely supported format for the object type?
- Will users have access to free viewers/software that support the proposed file type?
- What file size is produced and what are the implications for delivery to the user?
- Is the format easy to use?
- Will users require guidance or supporting documentation to allow them to access/use the objects?
- Does the organisation have separate user communities with different requirements for access?

See also [File formats and standards](#) for details of common preservation and access formats.

Legal issues for access

There are a variety of different legal issues that will probably need to be addressed when providing access to digital objects that will affect both the technological solutions that are deployed as well as who can access the material and when. This is one of the main access considerations that overlaps with acquisitions, as mentioned above, and it is essential that the correct information is gathered at that time to facilitate access requirements later in the life cycle. Without this information it may not be possible to properly manage access and may open the organisation to a number of potential legal risks.

Legal issues to be considered will include:

- Restrictions of use relating sensitivity and data protection
- Agreed embargoes on content where early access may represent a breach of contract
- Management of intellectual property rights, e.g. copyright

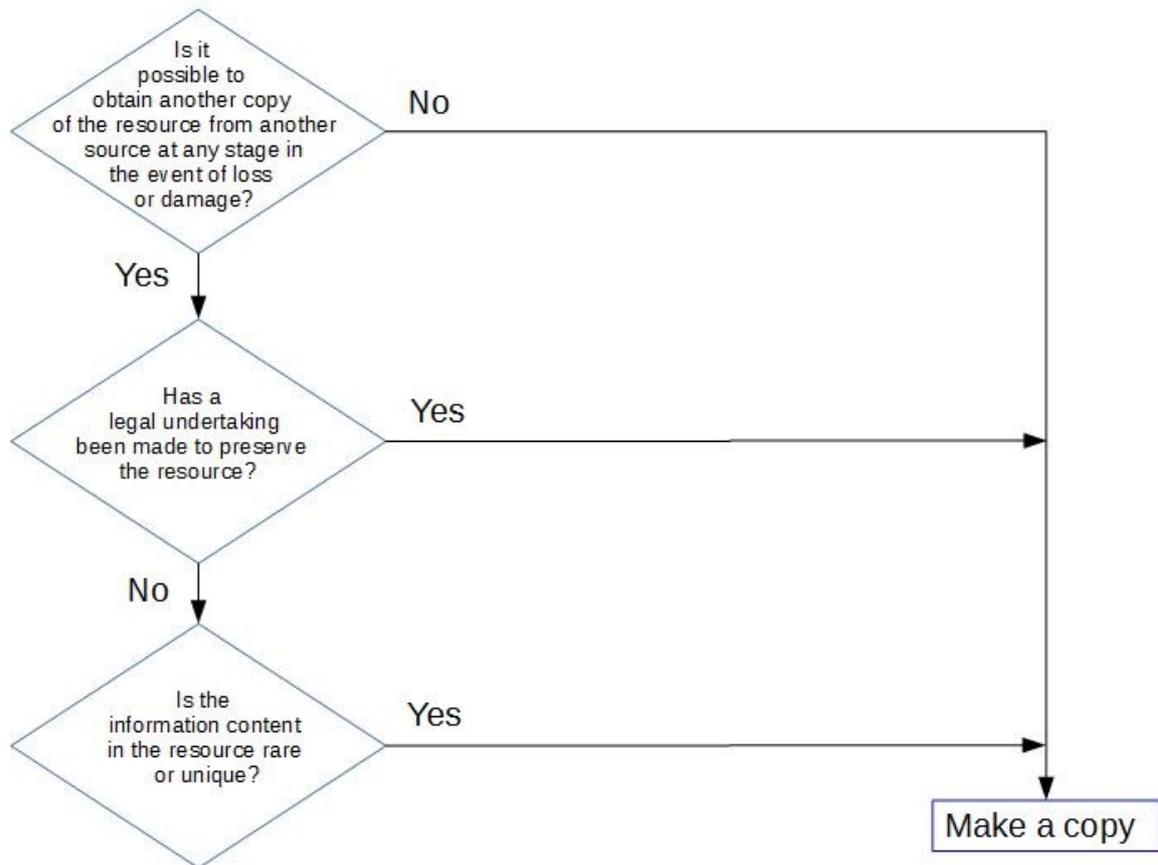
Management of IPR, in particular, should be aligned with the acquisition process with careful consideration given to transfer and ownership agreements and copyright licences put in place at that time. Licences must clearly state permitted access and reuse permissions, including third party licensing. These must then be clearly represented in policy and procedures for access, whether managed through a rights management system or by other methods.

Forms of access provision

The final key decisions an organisation must make are in the form of:

- Policy
- Procedure
- Free or charged services
- Online/Offline access, and the access environment provided
- Access for the disabled
- Storage and security

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. If this approach is taken a risk assessment needs to be undertaken consisting of some of the following questions (See also [Acquisition and appraisal](#) and [Storage](#)):



Conclusions

Access is closely linked to many other digital preservation issues and technologies covered in the Handbook. In particular you may wish to look at [Institutional policies and strategies](#), [Legal compliance](#), [Metadata and documentation](#), [Acquisition and appraisal](#), [Storage](#), [Legacy media](#), [File formats and standards](#), and [Information security](#).

Resources



Born-Digital Access in Archival Repositories: Mapping the Current Landscape, Preliminary Report August 2015

<https://docs.google.com/document/d/15v3Z6fFNydrXcGfGWXA4xzyWlivirfUXhHogqVDBtUg/preview?sle=true#>

This interesting document represents preliminary findings and analysis of a study and survey on current born-digital access practices in over 200 cultural heritage institutions. Respondents were primarily from the USA.

Reference model for an open archival information system (OAIS): Recommended practice (CCSDS 650.0-M-2: Magenta Book), Consultative Committee for Space Data Systems 2012

<http://public.ccsds.org/publications/archive/650x0m2.pdf>

This was later published as ISO 14721:2012, Space Data and Information Transfer Systems – Open Archival Information System (OAIS) – Reference Model, 2nd edition. The Access function within OAIS manages the processes and services by which consumers – and especially the Designated Community – locate, request, and receive delivery of items residing in the OAIS's archival storage. As such, it is the primary mechanism by which the archive meets its responsibility to make its information available to its user community. (135 pages).

Adrian Brown 2013 Practical Digital Preservation a how-to guide for organizations of any size

Chapter 9 (28 pages) of this book is devoted to the topic of providing access to users.



Community Owned digital Preservation Tool Registry COPTR

<http://www.digipres.org/tools/>

There are a large number of tools for access or that have access functionality incorporated in them. The Handbook recommends searching for them via the POWRR Grid tool within COPTR. The POWRR Tool Grid provides a set of interactive views designed to help practitioners identify and select tools that they need to solve digital preservation challenges. The Access, Use and Reuse column of the Grid identifies access tools for specific types of content or generic tools and systems that have access functions. Everything in the Grid is hyperlinked, so simply click through the displays until you find the information you are looking for. Clicking on the name of a specific preservation tool will reveal more detail on the COPTR wiki, which is where you should go to expand or update the tools information.

AIMS Born-Digital Collections: An Inter-Institutional Model for Stewardship. University of Hull, Stanford University, University of Virginia, and Yale University (2012)

http://dcs.library.virginia.edu/files/2013/02/AIMS_final.pdf

The AIMS (An Inter-Institutional Model for Stewardship) framework is a methodology for stewarding born-digital materials. It is divided into four main sections for high-level best practices for born-digital workflows: collection development, accessioning, arrangement and description, and discovery and access. Access primarily focuses on redaction and sensitive information. The appendices include, for example, sample processing workflow diagrams, an analysis of tools, and donor surveys. (195 pages).

Case studies



TNA case studies: Online access

<http://www.nationalarchives.gov.uk/archives-sector/online-access.htm>

a series of nine case studies published by TNA on how collections have been made more accessible by putting records online. They are drawn from a wide variety of archives.

Codebreakers: makers of modern genetics

<http://www.digirati.co.uk/case-studies/wellcome/>

A case study by digirati, the developers of the Wellcome Trust Library player focussing on the player its use in accessing the Francis Crick collection. The Wellcome Library's digital player is freely available for anyone to download and use. The player can be used to display all types of digital content, including cover-to-cover books, archives, works of art, videos and audio files. The software can be downloaded from the Wellcome Library GitHub account (<https://github.com/wellcomelibrary/player>).

Managing Risk with a Virtual Reading Room: Two Born Digital Projects, Michelle Light

http://digitalscholarship.unlv.edu/cgi/viewcontent.cgi?article=1462&context=lib_articles

Between 2010 and 2013, the University of California, Irvine, launched a site to provide online access to the personal papers of Richard Rorty and Mark Poster in the form of a virtual reading room. The virtual reading room mitigated the risks involved in providing this kind of access to personal, archival materials with privacy and copyright issues by limiting the number of qualified users and by limiting the discoverability of full-text content on the open web. The case study goes through each phase of research and thinking, including comparable projects happening at other institutions and lessons learned in a very open and informative way.

From Accession to Access: A Born-Digital Materials Case Study, by Cyndi Shein Journal of Western Archives Volume 5 Issue 1 (2014): 1-42

<http://digitalcommons.usu.edu/cgi/viewcontent.cgi?article=1036&context=westernarchives>

Between 2011 and 2013 the Getty Institutional Records and Archives made its first foray into the comprehensive ingest, arrangement, description, and delivery of unique born-digital material when it received oral history interviews generated by some of the Pacific Standard Time: Art in L.A. project partners. This case study touches upon the challenges and affordances inherent to this hybrid collection of audiovisual recordings, digital mixed-media files, and analog transcripts. It describes the Archives' efforts to develop a basic processing workflow that applies the resource-management strategy commonly known as "MPLP" in a digital environment, while striving to safeguard the integrity and authenticity of the files, adhere to professional standards, and uphold fundamental archival principles. The study describes the resulting workflow and highlights a few of the inexpensive technologies that were successfully employed to automate or expedite steps in the processing of content that was transferred via easily-accessible media and consisted of current file formats.

Metadata and documentation



Illustration by Jørgen Stamp digitalbevaring.dk CC BY 2.5 Denmark

Introduction

This section provides a brief novice to intermediate level overview of metadata and documentation, with a focus on the [PREMIS](#) digital preservation metadata standard. It draws on the 2nd edition of the DPC Technology Watch Report on Preservation Metadata. The report itself discusses a wider range of issues and practice in greater depth with extensive further reading and advice ([Gartner and Lavoie, 2013](#)). It is recommended to readers who need a more advanced level briefing.

Metadata is data about a digital resource that is stored in a structured form suitable for machine processing. It serves many purposes in long-term preservation, providing a record of activities that have been performed upon the digital material and a basis on which future decisions on preservation activities can be made in the future, as well as supporting discovery and use. The information contained within a metadata record often encompasses a range of topics. There is no clear line between what is preservation metadata and what is not, but ultimately the purpose of preservation metadata is to support the goals of long-term digital preservation, which are to maintain the availability, identity, persistence, renderability, understandability, and authenticity of digital objects over long periods of time.

Documentation is the information (such as software manuals, survey designs, and user guides) provided by a creator and the repository that supplements the metadata and provides enough information to enable the resource's use by others. It is often the only material providing insight into how a digital resource was created, manipulated, managed and used by its creator and it is often the key to others to make informed use of the resource.

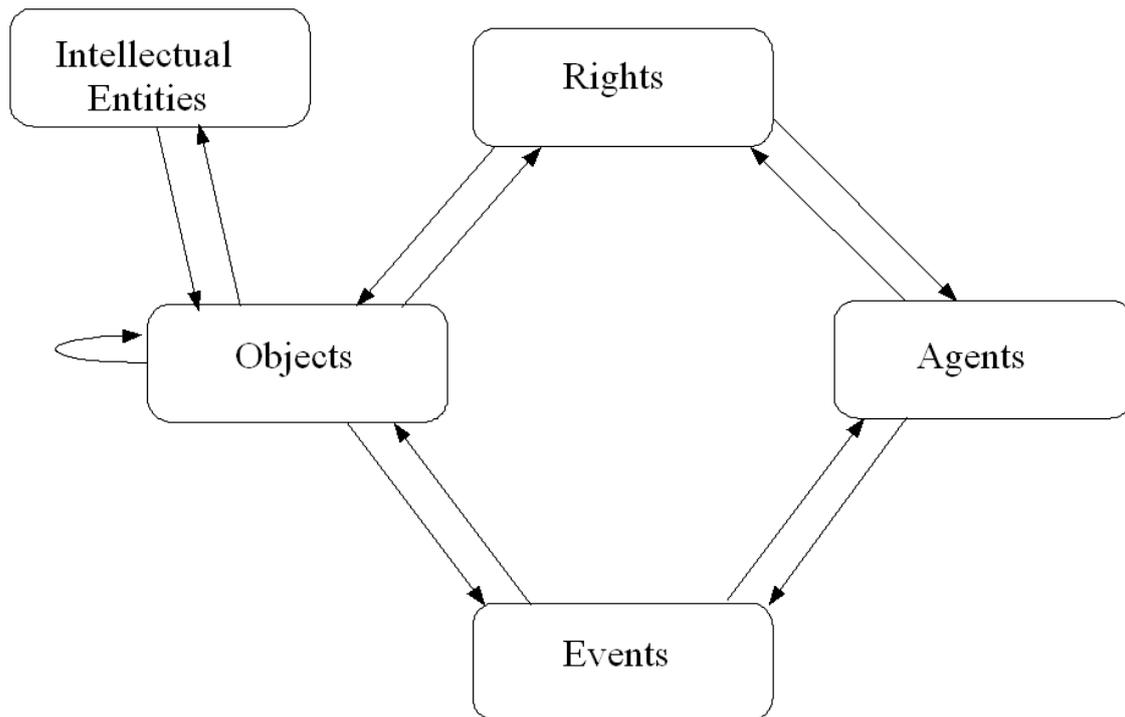
There are a number of factors which make metadata and documentation particularly critical for the continued viability of digital materials and they relate to fundamental differences between traditional and digital resources:

- **Technology.** Digital resources are dependent on hardware and software to render them intelligible. Technical requirements need to be recorded so that decisions on appropriate preservation and access strategies may be made.
- **Change.** While traditional materials may be preserved by predominantly passive preventive preservation programmes, digital materials will be subject to repeated actions, and there will be many different operators and quite possibly different institutions influencing the management of digital materials over a prolonged period of time. Recording actions taken on a resource and changes occurring as a result will provide a key to future managers and users of the resource.
- **Authenticity.** Metadata and documentation may be the major, if not the only, means of reliably establishing the authenticity of material following changes.
- **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) technologies have been used to control rights management and what (if any) implications there are for controlling access.
- **Future re-use.** It may not be possible for others to use the material without adequate documentation.
- **Cost.** It is expensive to create metadata manually and preservation metadata may not always be easily generated automatically. Additional metadata for digital preservation needs therefore requires careful cost/benefit trade-offs.

The PREMIS (PREservation Metadata: Implementation Strategies) Standard

[PREMIS](#) (PREservation Metadata: Implementation Strategies) is the international standard for metadata to support the preservation of digital objects and ensure their long-term usability. Developed by an international team, PREMIS is implemented in digital preservation projects around the world, and support for PREMIS is incorporated into a number of commercial and open-source digital preservation tools and systems.

The PREMIS Data Dictionary ([PREMIS, 2013](#)) is organized around a data model consisting of five entities associated with the digital preservation process:



1. **Intellectual Entity** - a coherent set of content that is described as a unit: e.g., a book
2. **Object** - a discrete unit of information in digital form, e.g., a PDF file
3. **Event** - a preservation action, e.g., ingest of the PDF file into the repository
4. **Agent** - a person, organization, or software program associated with an Event, e.g., the publisher of a PDF file
5. **Rights** - one or more permissions pertaining to an Object, e.g., permission to make copies of the PDF file for preservation purposes

Taken together, the semantic units defined in the PREMIS Data Dictionary represent the 'core' information needed to support digital preservation activities in most repository contexts. However, the concept of 'core' in regard to PREMIS is loosely defined: not all of the semantic units are considered mandatory in all situations, and some are optional in all situations. The **Data Dictionary** attempts to strike a balance between recognizing that there will be a significant overlap of metadata requirements across different repository contexts, while at the same time acknowledging that all contexts are different in some way, and therefore their respective metadata requirements will rarely be exactly the same.

Implementation

Although the PREMIS Data Dictionary is not a formal standard, in the sense of being managed by a recognized standards agency, it has achieved the status of the accepted standard for preservation metadata in the digital preservation community. A strength but also a limitation of the PREMIS Data Dictionary is that it must be tailored to meet the requirements of the specific context; it is not an off-the-shelf solution in the sense that an archive simply implements the Data Dictionary wholesale. Only a portion may be relevant in some digital preservation circumstances; alternatively, the

repository may find that additional information beyond what is defined in the Dictionary is needed to support their requirements. For example, the Data Dictionary makes no provisions for documenting information about a repository's business/policy dependencies, which may be needed to support preservation decision-making.

In short, each repository will need to invest some effort to adapt preservation metadata and documentation standards to its particular circumstances and requirements.

During implementation an 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 metadata and documentation for acquisitions and may notify and encourage suppliers or depositors to supply this information. Staff review and revise supplied information 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, information may be supplied by the third party in an agreed form which conforms to institution guidelines or in the supplier's native format.

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 validation or audit of the documentation and cataloguing for a resource shows this to be below a minimum acceptable standard.

A significant number of both users and suppliers of preservation metadata have adopted PREMIS and many of the initial obstacles to implementation have been addressed by them. The process of implementing PREMIS in a working environment is made easier by a number of tools which can extract metadata from digital objects and output PREMIS XML. The PREMIS Maintenance Activity maintains a webpage listing the most important tools available for use with PREMIS. It also includes an active email discussion list and a wiki for sharing documents. For further information see [Resources and case studies](#) below.

See also related sections of the Handbook including [Acquisition and appraisal](#), and [Preservation planning](#).

Resources



PREMIS Data Dictionary for Preservation Metadata, Version 3.0

<http://www.loc.gov/standards/premis/v3/index.html>

The PREMIS Data Dictionary and its supporting documentation is a comprehensive, practical resource for implementing preservation metadata in digital archiving systems. The Data Dictionary is built on a data model that defines five entities: Intellectual Entities, Objects, Events, Rights, and Agents. Each semantic unit defined in the Data Dictionary is a property of one of the entities in the data model. Version 3.0 was released in June 2015 (273 pages).

Preservation Metadata (2nd edition), DPC Technology Watch Report

<http://dx.doi.org/10.7207/twr13-03>

This report focuses on new developments in preservation metadata made possible by the emergence of PREMIS as a de facto international standard. It focuses on key implementation topics including revisions of the Data Dictionary; community outreach; packaging (with a focus on METS), tools, PREMIS implementations in digital preservation systems, and implementation resources. Published in 2013 (36 pages).



Tools for preservation metadata implementation

http://www.loc.gov/standards/premis/tools_for_premis.php

The PREMIS Maintenance Activity maintains a webpage listing the most important tools available for use with PREMIS. This contains entries on tools, in addition to pointers to others which may be used to generate METS (Metadata Encoding and Transmission Standard - an XML schema for packaging digital object metadata) files in conjunction with PREMIS. The majority of the tools listed are for extracting technical metadata from digital objects and converting it for encoding within the PREMIS Object entity. Others can be used for checking formats, or validating files against checksums



PREMIS website

<http://www.loc.gov/standards/premis/index.html>

The PREMIS Editorial Committee coordinates revisions and implementation of the PREMIS standard, which consists of the Data Dictionary, an XML schema, and supporting documentation. The PREMIS Implementers' Group forum, hosted by the PREMIS Maintenance Activity, includes an active email discussion list and a wiki for sharing documents. The wiki is a particularly useful resource for new implementers, as it includes materials from PREMIS tutorials, a collection of examples of PREMIS usage and links to information on PREMIS tools. The PREMIS Maintenance Activity maintains an active registry of PREMIS implementations.

Documenting your data

<http://www.data-archive.ac.uk/create-manage/document>

An excellent set of resources to assist researchers with the documentation and metadata for their research studies, drawn together by the UK Data Archive.

Archaeology Data Service Guidelines for Depositors

<http://archaeologydataservice.ac.uk/advice/guidelinesForDepositors>

The ADS Guidelines for Depositors provide guidance on how to correctly prepare data and compile metadata for deposition with ADS and describe the ways in which data can be deposited. There is also a series of shorter summary worksheets and checklists covering: data management; selection and retention; preferred file formats and metadata. Other resources for the use of potential depositors include a series of Guides to Good Practice, which complement the ADS Guidelines and provide more detailed information on specific data types.

Case studies



DPC case note: British Library ASR2 using METS to keep data and metadata together for preservation

http://www.dpconline.org/component/docman/doc_download/474-casenoteasr2.pdf

This Jisc-funded case study examines the 'Archival Sound Recordings 2' project from the British Library, noting that one of the challenges for long term access to digitised content is to ensure that descriptive information and digitised content are not separated from each other. The British Library has used a standard called METS to prevent this. July 2010 (4 pages).

Designing Metadata for Long-Term Data Preservation:DataONE Case Study

https://www.asis.org/asist2010/proceedings/proceedings/ASIST_AM10/submissions/435_Final_Submission.pdf

A short description of how PREMIS was utilized to specify the requirements for preservation metadata for DataONE (Data Observation Network for Earth) science data. 2010 (2 pages).

Preservica Case Study: Q&A with Glen McAninch, Kentucky Department for Libraries and Archives

<http://preservica.com/resource/qa-glen-mcaninch-kentucky-department-libraries-archives/>

Glen McAninch discusses the Importance of Provenance, Context and Metadata in Preserving Digital Archival Records.

PREMIS Implementations Registry

<http://www.loc.gov/standards/premis/registry/index.php>

The PREMIS Maintenance Activity maintains an active registry of over 40 PREMIS implementations with details of the repository and its use of PREMIS. Although not formally case studies, entries have details of practical experience e.g., [Creating a digital repository at the Swedish National Archives using PREMIS](#).

References

Gartner, R. and Lavoie, B., 2013. Preservation Metadata (2nd edition), *DPC Technology Watch Report* 13-3 May 2013. Available: <http://dx.doi.org/10.7207/twr13-03>

PREMIS, 2013. Data Dictionary for Preservation Metadata, Version 3.0. Available: <http://www.loc.gov/standards/premis/v3/index.html>