

Summary ELAG 2007 Workshop “Preservation of digital content”

Content

The workshop on digital preservation will start with interaction on general topics such as risk assessment and will become more specific with the discussion on the OAIS implementation at BnF. The workshop will end with the creation of a questionnaire for digital preservation strategies.

Schedule workshop

Wednesday 9 May

10.35-11.55 and 14.25-15.15

Seamus Ross on DPE risk assessment, PLANETS testbed, DCC metadata extraction

15.35-16.35

Emmanuelle Bermes will do a discussion on the OAIS standard

Topics:

1. What are the pre-requisites for an OAIS implementation?

The OAIS implies a transfer of responsibilities between the archive and producers/consumers/management. How to express the negotiation between these entities? We propose to set up "services contracts" that clarify the commitments of the archive and the producer on preservation and access.

Objectives:

- define the goals and policies for preservation and access
- have a clear vision of the risks (select at-risk material, set up priorities)
- define what material will be ingested in the archive.

How can we use the PAIMAS standard in this regard?

2. What to preserve? The definition of "digital cultural heritage".

-> we use the OAIS to define the scope of the digital preservation project.

definition of producers and user communities

definition of responsibilities and missions

-> digitisation, legal deposit, web archives, acquired material, digital masters, records

management: what different kinds of issues are these materials raising? Shall we have the goal to preserve anything, whatever the price and conditions? Can we refuse to ingest some material in the archive?

Thursday 10 May

09.40-10.30 and 11.00 ±11.30

Emmanuelle Bermes on OAIS

Topics:

1. Implementing the OAIS Information model: discussion on the use of packaging metadata formats

The OAIS defines a very detailed (although very conceptual) information model in order to gather all the information regarding a digital object that will be needed for its preservation. It is not required in the standard to implement the information model through a packaging format, but it is widely accepted in various communities to do so. Some packaging metadata formats like METS, MPEG-21 and XFDU have emerged.

What are the pros and cons of using a packaging format?

Comparative study of packaging formats: which one to choose?

Is it possible to define a common model that will make all these packaging formats interoperable? Discussion on the ORE initiative?

2. Implementing the OAIS functional model: what are the "quick wins" and what are the biggest issues?

The middle part of the functional model (Ingest - Data Management - Storage - Access) raises little or no issue. The interaction of the archive with producers and consumers still has to be clearly formalised through policies, sometimes in a machine-readable way. Expert systems will be able to apply the policies (e.g. rights management).

On the organisational point of view, the articulation between Administration and Preservation Planning raises questions. These two entities are not at the same level as the others. We suggest that the standard could be enriched with an emphasis on the necessary balance between stability and innovation that will be required to manage digital material in the long term.

11.30-12.00 and 15.30-16.50

Caroline van Wijk on digital preservation strategies

See discussion paper part 3

Discussion paper ELAG 2007

Workshop Preservation of Digital Content (third part of the workshop)
University of Barcelona, Spain, 9-11 May 2007

Caroline van Wijk
Koninklijke Bibliotheek, the National Library of the Netherlands
Caroline.vanwijk@kb.nl

Workshop part 3: Digital Preservation Strategies

Introduction

Objective workshop part 3

The objective of this third part of the workshop is to raise awareness on the importance for institutions to start thinking early on about defining (technological) approaches to ensure digital preservation of their collections.

It is important to start thinking about digital preservation strategies as soon as possible, because digital preservation (strategy) starts already at the *creation* of digital collections [1].

For this part of the workshop the participants are asked to create a questionnaire. They will define the questions and a few example answers for the questionnaire by discussing specific issues related to digital preservation strategies and by using the experiences of their institutions. The questionnaire can be used for starting up a discussion about digital preservation strategies within the workshop participant's organisations or other institutions.

Scope workshop part 3

The scope of this workshop part 3 will be *digital preservation strategies*. The definition of digital preservation strategy is described concisely by the UK Data Archive as

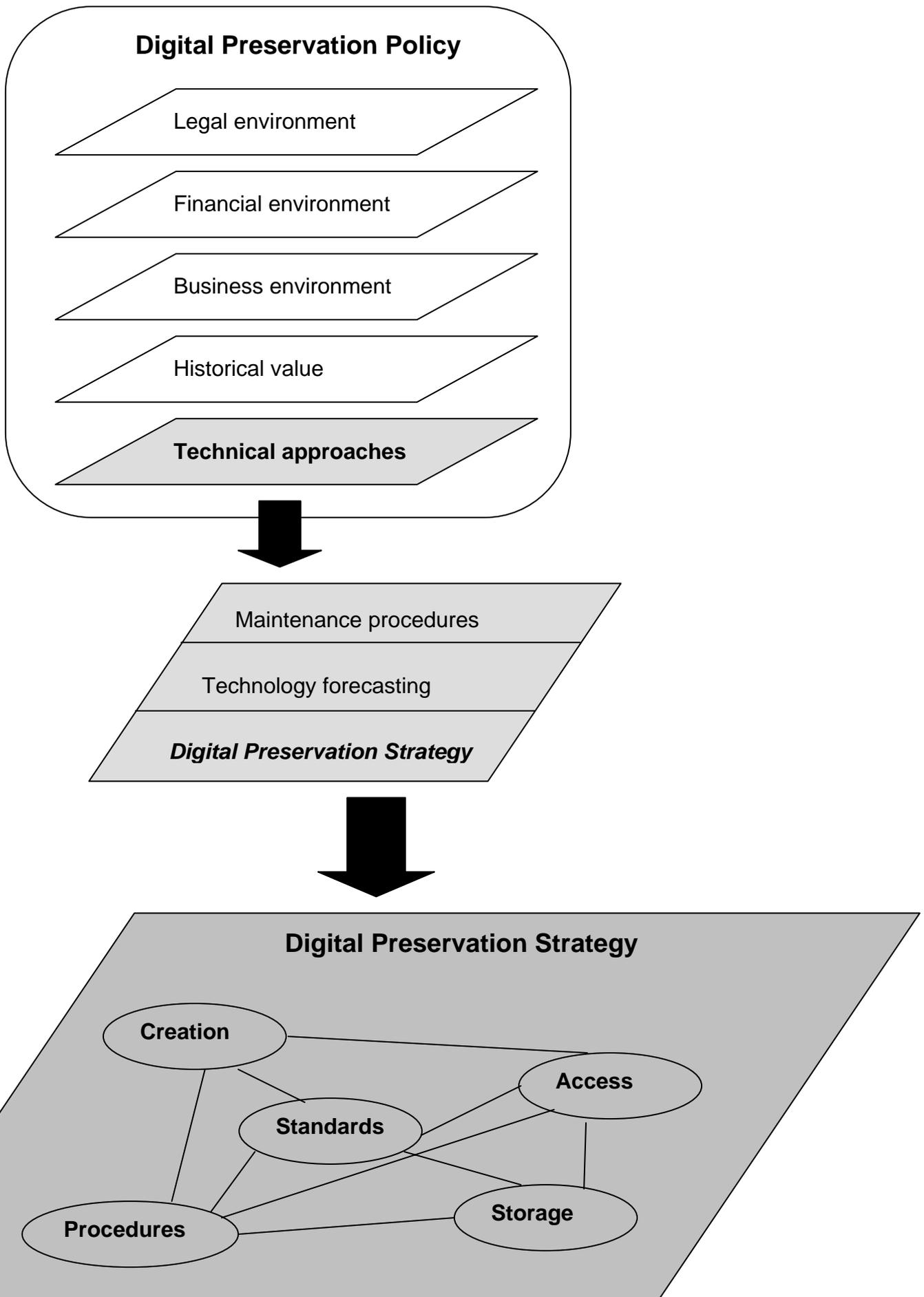
“Technical approach to ensure safe storage and permanent access to digital objects” [2]

Digital preservation policy - the framework around a digital preservation strategy within an organisation - will not be part of this part of the workshop. Although digital preservation policies and digital preservation strategies are entwined quite closely, the focus of this workshop will be on the questions (and answers) that should be asked to define the (technical) requirements that a preservation strategy should meet. Digital preservation policy has been the subject of workshops and publications, among which the work of ERPANET [3] (including a questionnaire for digital preservation policies!) and others.

The European project PLANETS [4] will also develop an organisational policy and strategy model that can be used for preservation planning. Examples of concrete policies are quite easy to find; many institutions have published their digital preservation policy on their website [5]. Also the first two parts of the Preserving Digital Content workshop touched on many subjects that are policy related.

This being said, it is important to understand the position of a digital preservation strategy within an organisation and its relation towards organisational, financial and legal aspects. A digital preservation strategy is to be embedded within the digital preservation policy of an institution and is influenced by the organisational, legal and financial environment in the institution. The context of a digital preservation strategy is visualised in figure 1.

Figure 1



The technical approaches within an organisation may also encompass maintenance procedures and technology forecasting next to a digital preservation strategy and are presented in the middle of the drawing. The former two approaches will also not be the focus of the workshop. The lower part of figure 1 presents the focus of the workshop – digital preservation strategy – in more detail.

A digital preservation strategy may comprehend an approach for creation of, access to and storage of digital objects and standards and procedures [1]. These sections are all related to each other and are also influenced by the institution's policy, which makes it difficult to start thinking about a digital preservation strategy. The workshop can help in this process by the discussion of the sections and by the creation of a (first draft of a) questionnaire that can be used outside of the workshop. The subdivision of a strategy in the sections *creation*, *access*, *standards*, *storage* and *procedures* will be used for the development of the questionnaire in the workshop.

The questionnaire on digital preservation strategies will contain:

- High level sections to categorise the questions and example answers: creation, storage, access, standards, procedures – provided by workshop leader
- Questions - product of the discussions by the participants of the workshop
- 'Example' answers to the questions that have been discussed and defined by the participants of the workshop using their experiences. As digital preservation strategies are embedded in an organisation's policy and influenced by that policy, various 'example' answers can be defined for the questions. Questions on digital preservation strategies do not have one correct answer.

Background information workshop

Questionnaire

The questionnaire sections are based on the aspects a digital preservation strategy can comprehend according to JISC [1]. The experiences at the National Library of the Netherlands and other institutions have been used for the brief introduction and explanation of the questionnaire sections where applicable.

Creation

During the last decade the number of digitisation projects has risen at the National Library of the Netherlands (KB). Digitisation guidelines have been developed to make sure that the created digital master files meet the requirements the digital preservation department has set for metadata and technical matters such as the use of specific file formats (non compression or lossless compression options). The KB is the creator of these digital objects and can influence the creation quite easily. However, the KB also archives millions of publications that are deposited by large publishers. The deposit contracts with the publishers contain several technical agreements (e.g. the file format of the publications). Also, guidelines for the creation of publications in the Portable Document Format have been created.

Currently, a file format risk assessment is being executed for specific file formats that have been deposited at the KB. This assessment will also produce a 'measurable' way

of assessing file format risks and it will be the starting point for specific definitions of digital preservation strategies.

Discussion keywords:

- Guidelines / selective deposit
- Archival period
- Preservation levels
- File format risk assessment
- Open standard formats, proprietary formats
- Compression, encryption, drm
- Metadata: significant properties
- Integrity checks

Storage

Many institutions have implemented or are implementing a digital archive at the moment. The choice for the type of storage medium in the archive is influenced by many aspects from the digital preservation policy of an institution, but also by aspects within the preservation strategy. For example, the requirements for access are important in the decision to store digital objects on tape or on hard disk. Publications in the e-Depot [6] – the digital archive of the KB – are stored on optical disks to ensure reasonable response times when users request a publication via the portal website. Digital master files from digitisation projects will be stored on tape as the response time for this type of digital object is not a priority. The storage of digital objects on CD-rom or DVD is convenient for distribution purposes, but is not considered useful for long-term archival storage.

Discussion keywords:

- Digital archive / document management system / file system
- Size of collections / type of objects
- Storage medium – refreshment cycle
- Metadata updates or new metadata
- Access requirements

Access

The manner of dissemination and display of digital objects will be highly related to an institution's policy and its contractual agreements with suppliers or depositors of the digital collections. One question could be whether an institution chooses to have separate archival and access copies of a specific collection.

An important part of digital preservation is to ensure permanent access to digital objects. The constant development of technology endangers the accessibility of digital objects. Preservation activities such as emulation and migration are based on different concepts. For example, emulation [7] and technology conservation strive to present the original digital object in its original technical environment– although both strategies try to meet this requirement by very different means. Migration [8] – and to some extent the Universal Virtual Computer [9] – are based on transformation of digital objects to current or (easier) accessible technical environments.

Before a specific preservation activity can be chosen for a collection, it is necessary to specify what characteristics of digital objects should be accessed in the future or which options an institution would like to present to the future user. Knowledge

about the collection and the significant properties of the objects in the collection and a definition of future user requirements are basic needs.

Most institutions – among which the KB – realise that all possible preservation strategies should be considered. Digital preservation is not ensured by one specific strategy only.

Discussion keywords:

- Dissemination / display
- Preservation level
- Focus on adaptation object or adaptation environment
- Technology approaches to ensure permanent access:
 - Migration (newer version, standard format)
 - Emulation (hardware, software)
 - Virtualisation
 - UVC
 - Do nothing
 - Technology conservation
 - Digital archaeology
- Future user expectations: significant properties

Standards

Standards are part of the creation, access, storage and procedure sections of a preservation strategy. Creation of and (permanent) access to digital objects deal for example with file formats that are de jure or de facto standards or not standard formats at all. The creation and maintenance of metadata involves standards that are familiar to anyone in the field of digital preservation: PREMIS [10], NISO [11] etc. The OAIS reference model [12] has been the basis for many digital archives. For example, the e-Depot at the KB is based on the OAIS Reference Model and is used for safe storage of digital objects.

PDF is a widely used proprietary file format and can be seen as a de facto standard. The KB archives millions of digital objects in this format as obsolescence or inaccessibility of PDF is not expected for a long time. PDF/A [13] is an ISO standard with open specifications and will be used for the normalisation of text based documents in formats that have closed specifications such as Word or WordPerfect.

Discussion keywords:

- OAIS
- Metadata: PREMIS, NISO, METS, MPEG21-DIDL
- de jure (open) standard formats
- de facto standard formats

Procedures

Procedures are involved in the creation, access to and storage of digital objects. Specific quality control procedures should be defined and executed in the operational workflow.

Procedures for preservation activities need to incorporate where the information about the (threatening) obsolescence of specific file formats will come from and they should define what sources of information will trigger the activities. File format registries such as GDFR [14] or PRONOM [15] could be the source of this information.

Discussion keywords:

- Quality control
- Network of digital preservation knowledge (PLANETS, GDFR)
- Periodical technical risk assessment
- Checksums
- Test procedures preservation activities such as migration

References

[1] JISC (2006) About Digital Preservation briefing paper JISC website.

http://www.jisc.ac.uk/publications/publications/pub_digipreservationbp.aspx

[2] UK Data Archive (2005) UK Data Archive Preservation Policy.

<http://www.data-archive.ac.uk/news/publications/UKDAPreservationPolicy0905.pdf>

[3] ERPANET publications:

ERPANET (2003) Digital Preservation Policy Tool.

<http://www.erpanet.org/guidance/docs/ERPANETPolicyTool.pdf>

ERPANET (2003) Selecting Technologies Tool.

http://www.erpanet.org/guidance/docs/ERPANETSelect_Techno.pdf

ERPANET, Policies for Digital Preservation: Seminar Report, ERPANET Training Seminar in Paris, January 29-30, 2003,

http://www.erpanet.org/events/2003/paris/ERPAttraining-Paris_Report.pdf

ERPANET (2002) Case Study Version 3.2

<http://www.erpanet.org/studies/docs/ERPANETCaseStudyQuestionsfinal.pdf>

Goethals, A. Digital Preservation Policies: Technical Considerations, SAA Program, August 6, 2004.

http://www.fcla.edu/digitalArchive/pdfs/tech_considerations_handout.pdf

[4] Preservation and Long-term Access through NETworked Services

The **PLANETS** project brings together European National Libraries and Archives, leading research institutions, and technology companies to address the challenge of preserving access to digital cultural and scientific knowledge. The four year project is funded by the European Commission Information Science and Technologies Framework Programme 6 Call 5 (FP6 Call 5).

<http://www.planets-project.eu/>

[5] Non comprehensive list of digital preservation policies:

National Library of Australia (Australia), <http://www.nla.gov.au/policy/digpres.html>

National Archives of Australia (Australia),

http://www.naa.gov.au/recordkeeping/er/digital_preservation/Green_Paper.pdf

British Library (UK),

<http://www.bl.uk/about/collectioncare/pdf/bldppolicy1102.pdf>

- UK Data Archive (UK),
<http://www.data-archive.ac.uk/news/publications/UKDAPreservationPolicy0905.pdf>
- Florida Digital Archive (USA),
<http://www.fcla.edu/digitalArchive/pdfs/DigitalArchivePolicyGuide.pdf>
- Cornell University Library (USA), <http://commondepository.library.cornell.edu/cul-dp-framework.pdf>
- [6] Information and publications on the e-Depot
<http://www.kb.nl/dnp/e-depot/dm/dm.html>
- Oltmans, E. & Lemmen, A. (2006) The e-Depot at the National Library of the Netherlands, *Serials*, 19 (1).
http://www.kb.nl/hrd/dd/dd_links_en_publicaties/publicaties/Serialsmarch2006.pdf
- [7] Modular emulation as a long-term preservation strategy for digital objects, J.R. van der Hoeven & H.N. van Wijngaarden, International Web Archiving Workshop 2005 (IWAW'05), Wenen, Austria, 2005.
<http://www.iwaw.net/05/papers/iwaw05-hoeven.pdf>
- [8] Migration research at the KB.
http://www.kb.nl/hrd/dd/dd_projecten/projecten_migratie-en.html
- [9] UVC for images
http://www.kb.nl/hrd/dd/dd_onderzoek/uvc_voor_images-en.html
- [10] PREMIS
<http://www.loc.gov/standards/premis/>
- [11] NISO Z39.87 Data Dictionary for Digital Still Images
http://www.niso.org/committees/committee_au.html
- [12] Reference Model for an Open Archival Information System
<http://public.ccsds.org/publications/archive/650x0b1.pdf>
- [13] PDF/A <http://www.pdfa.org/doku.php?id=pdfa:en>
- [14] Global Digital Format Registry
<http://hul.harvard.edu/gdfr/>
- [15] PRONOM
<http://www.nationalarchives.gov.uk/pronom/>

Glossary of terms

Term	Description
Digital preservation	Digital preservation is a series of managed activities necessary to ensure permanent access to digital objects for the long term.
Digital preservation policy	A digital preservation policy specifies the financial, legal and organisational environment for digital preservation within an institution.
Digital preservation strategy	Technical approach to ensure safe storage and permanent access

Term	Description
Emulation	to digital objects. The use of modern hardware and software to recreate an old computing environment and run old, obsolete files. An emulator mimics the old computing environment and allows the old software to run and access the file.
Migration	Migration is the transformation of digital data from one hardware/software environment to another to ensure permanent access for future use.
Digital object	Digital materials, both digitised and 'born digital' PREMIS specifies that a digital object can be a bit stream, file stream, file and representation.