

---

# Migration, Storage and Preservation

IMPACT Best Practice Guide

IMPACT project

Jeanna Nikolov-Ramírez Gaviria, Department for Research  
and Development at the Austrian National Library  
Niall Anderson, British Library

Released under Creative Commons - Attribution-NonCommercial-ShareAlike v3 Unported  
(International)

## Table of Contents

Executive Summary .....	1
Migration (including Refreshing, Maintenance, Emulation, Technology Preservation) .....	1
Long Term Preservation .....	2
Further reading on long-term preservation: .....	3

Best Practice Guide to Migration, Storage and Preservation

## Executive Summary

Even before a digital resource has been created, an institution must give consideration as to where it is to be stored, how it is to be stored, and for how long. While the cost of storing data digitally has been seen to fall considerably in recent years, the whole issue of digital storage is complicated by the potential obsolescence of the hardware and software that has been used to create the digital resource. This chapter offers a practical guide to the necessary considerations regarding storage and preservation, as well as the mitigation strategies most often used to keep digital resources accessible beyond the lifespan of the projects or technologies that created them.

## Migration (including Refreshing, Maintenance, Emulation, Technology Preservation)

To keep data accessible, current digital information needs to be transferred from one hardware and software setting to another and from one computer generation to subsequent generations. Migration can also be format-based, e.g. moving text files from an obsolete file format or to increase their functionality. The intervals of these migration actions will depend on available resources, urgency and changing technology lifecycles.

### Questions to consider:

- Which data should be migrated?
- Which format or data carrier to migrate to?
- At what intervals should data be migrated?
- What will be the agreed technical ‘triggers’ for a data migration?

Migration has storage implications for a large digitisation project, particularly where the project produces a master file and an access file. For an interactive guide to calculating the amount of storage needed throughout the lifecycle of a digitisation project, consult the IMPACT Storage Estimator (link [[https://www.surfgroepen.nl/sites/impactproject/oc/GA%20Annex%201%20Description%20of%20Work/OC2/OC2.1/IMPACT\\_Storage-Calculator\\_BSB.xls](https://www.surfgroepen.nl/sites/impactproject/oc/GA%20Annex%201%20Description%20of%20Work/OC2/OC2.1/IMPACT_Storage-Calculator_BSB.xls)]).

## Long Term Preservation

The high cost of digitisation and storage mean that the sustainability and longevity of content need to be considered at the onset of any mass-digitisation project.<sup>1</sup>

Current experience with the management of digitised content suggests that there are a number of general principles that need to be considered from a long-term preservation perspective:

- The longer-term viability of digital collections will be dependent on the quality and functionality of the digital objects created. Chapman and Kenney have argued for the creation of high-quality digital master files that include all of the features that are critical to the content's meaning.<sup>2</sup> They call this the "full informational capture" approach to digital conversion.
- In a digitisation context, preservation will primarily be focused on these digital master files, not with the derived versions created from them.
- The potential costs of ongoing storage and preservation need to be balanced against the potential costs of re-digitising content. While it may seem absurd to plan a large-scale digitisation project on the understanding that the process may have to be repeated at some stage in the future, the uncertain costs of long-term preservation as well as ongoing improvements in digitisation techniques may mean that this will ultimately be the most efficient or cost-effective approach. This strategy would naturally, however, depend on the retention and preservation of the original content objects.

There is a growing body of knowledge about the key issues that will need to be considered by those organisations that accept responsibility for the long-term stewardship of digital content.

Some of these primarily relate to the organisational and technological aspects of developing digital preservation services, initially the Reference Model for an Open Archival Information System (OAIS) (ISO 14721:2003), but now supported by a growing body of work on repository audit and certification.

There has also been a growing focus on the types of information that would be required to support the preservation of digital content over time. One strand of this focuses on the technical information on file formats and other types of "representation information" that are needed to support digital preservation systems, including registry services like PRONOM<sup>3</sup> together with associated format identification, validation and characterisation tools like DROID<sup>4</sup> or JHOVE2.<sup>5</sup> The other strand is centred on what is known as preservation metadata, currently focused on emerging standards like the PREMIS Data Dictionary and the NISO Z39.87 Data Dictionary - Technical Metadata for Digital Still Images.

The economics of digital preservation has also been gaining some attention in recent years. For example, the two-phase UK LIFE (Lifecycle Information For E-Literature) project has attempted to model digital

---

<sup>1</sup>Oya Y. Rieger, *Preservation in the age of large-scale digitisation*, Washington, D.C.: Council on Library and Information Resources, February 2008.: <http://www.clir.org/pubs/abstract/pub141abst.html> Retrieved 13.03.2011

<sup>2</sup>Stephen Chapman and Anne R. Kenney, "Digital conversion of research library materials: a case for full informational capture," *D-Lib Magazine*, October 1996. <http://www.dlib.org/dlib/october96/cornell/10chapman.html> Retrieved 13.03.2011

<sup>3</sup>The National Archives, PRONOM technical registry; 2011; <http://www.nationalarchives.gov.uk/pronom/> Retrieved 13.03.2011

<sup>4</sup>The National Archives, DROID (Digital Record Object Identification); 2011: <http://droid.sourceforge.net/wiki/index.php/Introduction> Retrieved 13.03.2011

<sup>5</sup>Stephen Abrams, Sheila Morrissey, and Tom Cramer, "The next generation JHOVE2 architecture for format-aware characterisation," Proceedings of iPres 2008: the Fifth International Conference on Preservation of Digital Objects, London, September 29-30, 2008, pp. 86-92.

preservation costs over five, ten and one hundred years and apply these models in detailed case studies. The international Blue Ribbon Task Force on Sustainable Digital Preservation and Access is currently analysing existing costs models and aims to develop a set of economically viable recommendations on the adoption of digital preservation strategies.<sup>6</sup>

Many cultural heritage organisations have now begun to consider their digital preservation requirements in more detail and these are beginning to be embodied in digital preservation policies and strategies. Cornell University Library has produced a good example of a strategy that includes digitised content.

**Questions to consider:**

- Where will the archive be stored (internally, off-site, or both)?
- Where and how will the archive be managed (internally or off-site)?
- Which storage media will be used?
- How much storage space is needed to store all of the digitised data, master files and surrogates
- How does digitised content fit into the organisation's wider digital preservation strategy?
- What baseline functionality needs to be embodied in the digital master?
- What preservation strategies might be adopted for the digital master?
- What implications might there be for metadata capture?
- Would it be more cost effective and efficient to periodically re-digitise content?
- What policies need to be adopted for derivative or surrogate objects?

For a guide to estimating storage costs please see the IMPACT Storage Estimator available to download from the IMPACT Website.

## Further reading on long-term preservation:

- **CASPAR (Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval)**

<http://www.casparpreserves.eu/>

A collaborative research project funded by the European Union as part of its Sixth Framework Programme investigating the development of OAI-based digital preservation tools and services.

- **Cornell University Library Digital Preservation Framework (December 2004)**

<http://hdl.handle.net/1813/11230>

This is a good example of a comprehensive institution-based policy for the preservation of all kinds of digital content. The policy gives priority to born digital materials and digitised materials with no print analogue, but states that "reasonable measures" will be taken to extend the life of those objects with a readily available print analogue. However, there is a caveat to the effect that the cost of redigitisation would need to be weighed against the cost of preserving the digital versions.

- **LIFE, LIFE2 and LIFE3 (Life Cycle Information for e-Literature)**

---

<sup>6</sup>Blue Ribbon Task Force on Sustainable Digital Preservation and Access homepage: <http://brtf.sdsc.edu/> Retrieved 13.03.2011

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

A three-phase UK project funded by the Joint Information Systems Committee that has attempted to model digital preservation costs over five, ten and one hundred years. LIFE3 is scheduled to complete in August 2010.

- **NESTOR - Das Kompetenznetzwerk zur digitalen Langzeitarchivierung/Network of Expertise in Long-Term Storage and Availability of Digital Resources in Germany**

<http://www.langzeitarchivierung.de/>

Nestor is a collaboration of major German educational and cultural heritage organisations focused on the long-term stewardship of digital resources in Germany and on international collaboration. The nestor website gives access to a comprehensive collection of online materials on digital preservation.

- **OAIS (Reference Model for an Open Archival Information System)**

<http://public.ccsds.org/publications/archive/650x0b1.pdf>

The OAIS reference model (ISO 14721:2003) provides a general framework for long-term digital preservation and access. It includes definitions of relevant terms and can be used to help define or compare preservation architectures. It has also helped influence the design of many digital preservation systems, including the DIAS system used by the Koninklijke Bibliotheek.

- **PADI (Preserving Access to Digital Information)**

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

PADI (Preserving Access to Digital Information) is an annotated list of online resources relating to all aspects of digital preservation. The National Library of Australia maintains the service with the support of an international advisory group. There is a specific section on digitisation (<http://www.nla.gov.au/padi/topics/69.html>)

- **Open Planets Foundation**

<http://www.openplanetsfoundation.org/>

PLANETS was a collaborative research project funded by the European Union as part of its Sixth Framework Programme, building practical tools that will support long-term access to digital cultural and scientific assets. These include a tool to support preservation planning and a test bed environment that will help support the gathering of empirical data on characterisation services and preservation strategies. The Open PLANETS Foundation aims to build on the work of the earlier project and expand its reach to other organisation and centres of competence in digitisation and storage

- **TRAC (Trustworthy Repositories Audit and Certification)**

<http://www.crl.edu/content.asp?11=13&12=58&13=162&14=91>

The *Trustworthy Repositories Audit and Certification (TRAC) Criteria and Checklist* is just one of several emerging frameworks that can be used for the internal or external evaluation of digital preservation services. TRAC is based on a checklist originally developed by the Digital Repository Certification Task Force. Other approaches include **DRAMBORA** (Digital Repository Audit Method Based on Risk Assessment),<sup>7</sup> which is primarily designed to help support the self-assessment of digital preservation services.

---

<sup>7</sup>Digital Repository Audit Method Based on Risk Assessment (DRAMBORA): <http://www.repositoryaudit.eu/> Retrieved 13.03.2011