

Digital Preservation

Continued access to authentic digital assets

Briefing Paper

November 2006

Print materials can survive for centuries and even millennia without direct intervention. In contrast, digital materials may need active management and preservation in order to survive even a decade.

The increased use of digital technologies in UK education and research institutions has resulted in a massive growth in the volume of digital assets being created. Many of these assets have lasting value and must be preserved to ensure investments are maximised, knowledge can be reused, accountability is assured, and organisational memory is retained. This requires considerable input from a range of stakeholders into the technical, financial, organisational and cultural issues involved to ensure that authentic, meaningful and reusable resources are preserved.

This briefing paper offers an introduction to these issues. It defines digital preservation and its relationship with digital curation, illustrates some of the advantages in preserving resources created or stored digitally, outlines the policy and strategy requirements, and identifies the essential roles and responsibilities that a preservation strategy should address.

What is digital preservation?

Digital preservation is the series of actions and interventions required to ensure continued and reliable access to authentic digital objects for as long as they are deemed to be of value. This encompasses not just technical activities, but also all of the strategic and organisational considerations that relate to the survival and management of digital material.

Digital objects will cease to be accessible without active management and intervention. The biggest risk to the accessibility of digital objects is the continual development of computing hardware and software. Many digital files or formats are dependent upon a particular computing environment for accurate presentation of

their content. Any change to the rendering environment could result in change to the rendered representation of a resource (or result in not being able to render the resource at all). The severity and impact of the change varies considerably between objects or environments and can often have a detrimental effect on the authenticity and integrity of a resource. This in turn affects its reliability, trustworthiness and capacity for subsequent reuse. Planned and tested strategies to counter these risks are therefore vital.

Digital Curation

Digital curation is all about maintaining and adding value to a trusted body of digital information for future and current use; specifically, the active management and appraisal of data over the entire life cycle. Digital curation builds upon the underlying concepts of digital preservation whilst emphasising opportunities for added value and knowledge through annotation and continuing resource management. Preservation is a curation activity, although both are concerned with managing digital resources with no significant (or only controlled) changes over time.

The current rate of technological change may mean that preservation actions, such as migrating to more accessible or durable formats may be required after as little as five years. Digital preservation should therefore be addressed from as early in the object life cycle as possible, particularly as the manner in which a resource is created has a significant impact on its durability. This requires effort and input from an organisational and cultural perspective as well as a technical one.

Authenticity

Authenticity – the quality of being authentic – allows digital assets to be reliably reused. An authentic digital resource is one that is what it purports to be, is free from corruption, and is intact in all essential respects. Authenticity should be a consideration in all digital preservation activities.

Organisational infrastructure to support preservation

Proper preservation processes require support from organisational policy and a pro-active strategy. Both are necessary to ensure sufficient financial and organisational commitments to achieve successful digital preservation.

A digital information audit is a necessary precursor to the development of preservation policy and strategy. It should identify which resources are held, and which need to be preserved. Results from the information audit, particularly in relation to risk management and critical data, feed into the development of preservation policy.

A high-level policy document should explain why digital preservation is necessary and identify the areas where attention is needed. This can then form the basis of a more detailed and related strategy. It is important to maintain two separate, though related, documents, as the strategic implementation targets a different audience from the policy document and explores the issues at a far finer level of granularity. Both policy and strategy should be regularly assessed to ensure changes to the organisational, legal, or technical infrastructure are accounted for.

Digital preservation policies are most effective when integrated into the overall organisational policy framework, reflecting both organisational commitment to preservation and the importance afforded to the activity. This helps reinforce these perceptions amongst staff and users. Topics to consider in a digital preservation policy include, but are not limited to:

- Justification for preservation
- Organisational and financial commitment
- Preservation of authentic resources and quality control

- Metadata creation
- High-level identification of roles and responsibilities
- Training and education

Backups vs Preservation

Disaster recovery strategies and backup systems are not sufficient to ensure survival and access to authentic digital resources over time. A backup is a short-term data recovery solution following loss or corruption and is fundamentally different to an electronic preservation archive.

A well-constructed and written digital preservation strategy will provide further detail on implementation of the areas outlined in the policy and will identify and lead digital preservation activities across the life cycle of the resources. The strategy may take into account creation, storage, access, standards and procedures (particularly for quality control), and can explicitly determine appropriate preservation activities to combat technological obsolescence for different types of resources, such as migration or emulation (or derivatives of each).

The selected approach must be sure to preserve the resources in an authentic manner, without significant or unknown loss or change. A preliminary test and assessment of the viability of a given approach on a particular set of resources (such as migration of a set of text documents or a group of databases to a more durable and accessible format) is the most reliable way to achieve this.

Metadata is a significant component of preservation. Metadata is data about data; it is vital for assuring the context of a resource, for documenting its provenance and authenticity and for enabling resources to be discovered. Metadata can also document the policies and processes used by a preservation service, eg detailing what transformations have taken place and providing a secure audit trail. Addressing metadata requirements in a written strategy, such as identifying required elements and selecting appropriate schemas,

Digital preservation strategies

Migration

The conversion of data into current or more widely accessible formats. Can be implemented in a number of ways, including backwards compatibility, conversion to standards, and canonicalisation.

The advantage of this approach is that data are maintained in a currently accessible format. The main disadvantage is potential unknown data loss, exacerbated by the need for recursive migrations over time. The 'migration on request' strategy, a derivative of this approach, helps to address this problem by focusing on the transformation of migration tools rather than the objects themselves.

Emulation

The use of modern hardware and software to recreate an old computing environment and run old, obsolete files. The source file must be safely stored with the correct access software and an emulator written that mimics the old computing environment and allows the software to run and access the file. The main advantage of this approach is that error-introducing data conversions are avoided. The disadvantage is the technical complexity of the approach.

Technical strategies based on the production of non-digital backups (eg on paper or microfilm) or those reliant on museums of obsolete hardware are not generally considered to be acceptable approaches to preserving born-digital objects, except (perhaps) where particular circumstances necessitate this.

ensures that a consistent approach is developed and followed across the life cycle for all resource types.

Strategies will require frequent review to ensure the technical specifications remain compatible with emerging technologies.

Roles and responsibilities

Stakeholders in the digital preservation process occupy diverse roles, yet continuity and compatibility in their digital preservation responsibilities are required to ensure success.

Resource creators have a responsibility to create well-formed and sustainable resources, using open and standard file formats wherever possible. Guidance on achieving this is necessary, as good practice for sustainability is often absent in an environment where resources are created only for the 'here and now'.

Resource managers, who may also be resource creators, are responsible for ensuring information is properly managed and remains accessible whilst in their care. Insofar as preservation is concerned, this can involve everything from preservation planning for avoiding obsolescence, to developing strategy, taking

responsibility for preservation, and liaising with IT staff or external preservation service providers (including national bodies) who provide the technical tools and infrastructure for digital preservation. Resource managers can occupy a range of roles, including database or collection manager, records manager, archivist or librarian, IT manager, administrative staff, academics and researchers. A resource can have more than one resource manager during its lifetime, so specific activities will vary at different stages of the object life cycle. The fundamental responsibility is to ensure that activities are driven by the need to provide continued access to the resource(s) concerned.

Other parties not directly involved in institutional digital preservation nevertheless play a contributory role. These include senior management, funding bodies, advisory bodies, national libraries and archives, national and international digital preservation research initiatives, external preservation service providers, and even data reusers. External and national initiatives are particularly useful resources from which experiences and outputs can be harvested and applied. Digital preservation is a complex set of challenges for any institution to undertake alone. Sharing the challenge makes it easier to address.

JISC and Digital Preservation

Ensuring long-term preservation of and continuing access to scholarly and educational resources is an important strategic area for JISC. JISC has been a key driver in initiatives aimed at embedding digital preservation and curation within UK higher and further education institutions. JISC has undertaken various activities to help institutions address the challenges of digital preservation. A number of digital preservation-related programmes are contributing to this effort, such as the Supporting Digital Preservation and Asset Management in Institutions (2004–2006) programme and the establishment of the Digital Curation Centre (DCC). The DCC provides a national focus for research and development into curation issues and promotes expertise and good practice for the management of research outputs in digital format.

The forthcoming 2007–2009 JISC Strategy recognises the need for development of software and management tools to support digital preservation within the proposed national infrastructure of digital repositories. This is a priority and key deliverable in achieving JISC's aim 'to deliver innovative and sustainable ICT infrastructure, services and practice that support institutions in meeting their missions'. Digital preservation is identified as a crucial area for further development during the period 2006–2009 and JISC has committed additional funding to this area.

This paper has been written by Maureen Pennock, Digital Curation Centre Research Officer at the University of Bath, on behalf of JISC.

Alternative formats of the briefing paper can be found at:
www.jisc.ac.uk/publications

Further information and resources

The Open Archival Information System (OAIS)

Reference Model is an ISO standard (ISO 14721) that identifies the processes required for long-term preservation within an archival repository and establishes a common framework of terms and concepts. The full standard is available for purchase from ISO. An overview of ISO archiving standards is available at:
<http://nost.gsfc.nasa.gov/isoas/>

The PREMIS Data Dictionary builds upon the OAIS model and is the emerging standard for preservation metadata. The data dictionary and other PREMIS project outputs are available from:
www.oclc.org/research/projects/pmwg/default.htm

Digital Curation Centre

www.dcc.ac.uk

Digital Preservation Coalition

www.dpconline.org

PADI (Preserving Access to Digital Information)

www.nla.gov.au/padi

ERpanet (Electronic Resource Preservation and Access Network)

www.erpanet.org

Digital Preservation at the UK National Archives (TNA)

www.nationalarchives.gov.uk/preservation/digital.htm