



JISC Digital Media

www.jiscdigitalmedia.ac.uk

An Introduction to Digital Preservation

Last updated: **09 January 2009**

Published in: **Managing your digital resources**

Tags: **bce** | **digital preservation** | **metadata** | **quality assurance**

Summary

Digital preservation aims to ensure that a digital collection remains usable, regardless of the inevitable changes in technology the future will bring. Without the appropriate preservation methods in place a digital collection can easily become inaccessible and so useless in just a few years. This paper discusses the preservation of digital output and the associated metadata.

The problem

Digital data is under consistent threat of loss for a number of reasons. The data file type used, the media that the data is saved on, and the hardware and software systems needed to read the digital data can all points of vulnerability. But digital collections also come under threat from changes to organisational management and culture as well as from financial pressures and changes of financial priority.

At present, it can be argued that there is no digital media that is proved to be as durable and reliable as some analogue media. All digital media can and will deteriorate over time leading to loss of information. However, digital data can theoretically be reproduced infinitely, so the fragility of any piece of media is not necessarily as large a risk to the survival of digital information as another threat: technical obsolescence.

Typically, digital media will far outlive both the hardware and the software that support it. Technology is advancing so quickly that steps must be taken to ensure that data can be retrieved well into the future.

Approaches to preservation

The protection and long-term preservation of a project's digital output and metadata needs to be considered even before digitisation begins. Decisions made at the earliest stages of a project can and will have an impact on the effectiveness of the whole digital preservation strategy. It is particularly important that digitisation projects are fully documented as they progress. Full documentation of technical solutions and project delivery will give those undertaking the preservation strategy an understanding of how the project was conceived, developed and produced.

Strategies have to be put in place to guarantee that the collection survives through technological changes, ensuring its continued accessibility and usability. There are three common approaches to digital preservation:

- **Migration**

Migration describes the process of copying content from one format (such as a CD-ROM) onto a newer format (such as a solid state flash drive).

- **Refreshment**

A related process is refreshment. Refreshment involves copying data onto a newer example of the same format (such as from an old CD-ROM to a new CD-ROM).

- **Emulation**

Emulation is a more involved process of accessing data on a system other than the one it was made for. Commonly, this will be because an original system is no longer available. Playing vintage computer games on a contemporary games emulator is a good example.

Preservation in practice

Whichever approach or combination of approaches is chosen, it is often helpful to make a distinction between a 'master generation' of digital data and at least one surrogate 'delivery generation'. The master generation should contain as much intellectual, visual or audio content as possible and must be saved in a standard (non-proprietary) file format and it should preferably be duplicated across multiple locations. Delivery generations of data, however, may be re-sized, compressed, and saved in whichever format is suitable for delivery to the user. Delivery versions are typically of lower quality (more compressed) than their original master files. Defining the status and thereby the relative importance of a file helps immensely in the task of preservation.

Sustainable storage

A storage solution should be decided upon before producing any digital output, as it is of prime importance. Strategies for both online (hard drive) and offline storage (such as DVD discs) should be considered for the collection. Due to the large size of master files an entire collection can be very substantial in size, possibly requiring a mixed architecture for data storage. The size of both master files and any surrogates has implications for the amount of storage space required and should be calculated at the outset of the project. This factor may dictate the resolution at which materials are captured.

Master files can often be stored offline, since they are infrequently accessed. However, this does mean that automated error checking or metadata extraction tools cannot easily be used.

The delivery generation of data is in continual use and will typically be stored online.

A variety of digital storage media are available for offline storage, including CD-ROM, DVD-ROM, LTO (Linear Tape Open) and DLT (Digital Linear Tape). Hard drives too, can be used as 'plug and play' storage devices and stored safely on shelves away from computers.

However, as none of these have been around long enough to be proven as a viable long-term storage medium. It is best practise to write archival data to more than one type of media and then store these in different locations.

Regardless of the solution chosen for offline storage, media must always be handled appropriately and stored in the correct environmental conditions.

Online storage is often mirrored across multiple disks using redundant disk arrays (RAID).

A routine error-checking schedule should be implemented and a strategy drawn up for migrating data and metadata to suitable formats as necessary. If a file format is becoming obsolete and a migration is planned, archival master files should be migrated to new formats that are non-proprietary (such as TIFF for images, motion JPEG2000 for video or AIFF for audio) wherever possible. Quality control checks should follow any migration or refreshment so that any loss of data integrity can be identified and quickly addressed.

Responsibilities

Digital preservation is an active task. Whereas in the past, it was possible to put your assets in a box under the stairs and find them usable in ten years time - this is an unlikely scenario with digital media. It is imperative that the responsibility for all digital resources is firmly assured and known to all stakeholders. Digitisation projects should have, as part of their project specifications a policy which covers:

- Who the digital resource or collection belongs to and who is responsible for its upkeep
- What the process is for deciding when and how refreshment/migration takes place and who makes the decision
- Where the budget is coming from for this ongoing digital preservation investment

While it is clear that a technical strategy is necessary to ensure digital preservation, it is also important that the venture receives an organisational commitment.

There are costs associated with the preservation of a digital collection and the costs will vary according to how the process is undertaken. For example, organisations that undertake to maintain and preserve a digital collection in-house may face larger costs in terms of staff time and training than those that contract external organisations to undertake the process. However, if external contractors are used, digital preservation skills are unlikely to be developed in-house. It must be remembered that regardless of how digital preservation is undertaken, there will be associated costs.

Digitising materials requires a particularly large investment of both time and money. From an organisational perspective, a commitment from senior management to the longevity of the resource must be realised along with the release of an appropriate amount of resources in terms of money and other support if data is to be preserved.

Conclusion

It is interesting to note that due to technological obsolescence and media fragility many consider it possible that future generations will have less information about Gulf War conflicts (recorded on digital media) than the First World War (recorded on analogue media). The greatest asset of digital information - the ease with which it can be copied or transferred - is paralleled by the ease with which the information can be corrupted or deleted.

Further reading

A wealth of digital preservation related activity is now underway and a strong preservation community is developing. JISC (Joint Information Systems Committee) are heavily involved with digital preservation in the UK. Along with Charles Beagrie, JISC have produced the **Digital Preservation Policies Study** (<http://www.jisc.ac.uk/whatwedo/programmes/preservation/2008digprespolycystudy.aspx>) which looks at the long term institutional requirements for successful preservation of data. JISC also host the **digital preservation mailing list** (<https://www.jiscmail.ac.uk/cgi-bin/webadmin?A0=digital-preservation>) . While the Digital Preservation Coalition provide advocacy for digital preservation activities and often host related events and conferences.

Internationally too, there is much activity relating to digital preservation. The **PLANETS** (<http://www.planets-project.eu/>) project brings together European libraries and archives in order to best preserve scientific and cultural data. The Library of Congress in the US undertake preservation research and publish a **Digital Preservation Newsletter** (<http://www.digitalpreservation.gov/>) .

iPRES (<http://www.bl.uk/ipres2008/>) is a series of conferences dealing with digital preservation related topics from strategy to implementation, and from international and regional initiatives to small organisations.

In addition JISC Digital Media provide a range of advice documents covering topics such as choosing sustainable file formats or digitising specific types of material.

Last updated: **09 January 2009**

Published in: **Managing your digital resources**

Tags: **bce** | **digital preservation** | **metadata** | **quality assurance**

Copyright © 2009 JISC Digital Media | A JISC Advisory Service

